Protocol Design for Machine-to-Machine Networks

Aijaz, Adnan; Aijaz, Adnan

Awarding institution:
King's College London

Protocol Design for Machine-to-Machine Networks

by

Adnan Aijaz

A Thesis Submitted for the Degree of

Doctor of Philosophy

at

King's College London

October 2014

To Mom, Dad, Hamid, and Lubna

Acknowledgements

First and foremost, I thank Allah Almighty for giving me the strength and the means to complete my doctoral studies.

This wonderful and challenging journey, which started in Oct. 2011, would not have been possible without the guidance and support of several individuals. I owe my deepest gratitude to Prof. Hamid Aghvami for his continuous support, relentless and insightful guidance, and inspirational mentorship. Thank you Sir, for everything you did for me and for always believing in me. Your seemingly simplistic but actually profound rule of *3 Selfs* for success will stay with me forever.

Special thanks to Dr. Reza Nakhai, Dr. Xiaoli Chu, and Dr. Vasilis Friderikos for their helpful guidance and useful advice related to my research. I would also like to thank Prof. Mischa Dohler for appreciating my research.

I would also like to thank my examiners: Prof. Kin K. Leung (Imperial College London) and Dr. John Mitchell (University College London), for making my viva an enjoyable experience as well as for their valuable suggestions which have led to significant improvements in the presentation and quality of this thesis.

I am particularly grateful to all my current and erstwhile colleagues at CTR, including Paul, Oliver, Auon, Diogo, Ana, Tuan, Alex, Jing, Mati, Shuyu, Menglan, Hong, Ehsan, Omar, Christoforos, Giorgos, the visiting researchers (Massimo, Baldomero, and Zainab), and Yaqubi (the wizard). Thank you guys for all the support and cooperation you have extended to me and for making my stay at CTR as the most memorable time of my life.

I would also like to acknowledge the funding support of the School of Natural and Mathematical Sciences, King's College London.

Last, but not the least, I wish to thank my family, who has always stood by me through thick and thin. It is to them that I owe all the success in life.

Abstract

Machine-to-Machine (M2M) communications is an emerging communication paradigm that provides ubiquitous connectivity between devices along with an ability to communicate autonomously without human intervention. M2M communications acts as an enabling technology for the practical realization of Internet-of-Things (IoT). However, M2M communications differs from conventional Human-to-Human (H2H) communications due to its unique features such as massive number of connected devices, small data transmissions, little or no mobility, requirements of high energy efficiency and reliability, etc. These features create various challenges for existing communication networks which are primarily optimized for H2H communications. Therefore, novel solutions are required to meet the key requirements of M2M communications. In addition, enhancements are required at different layers of the protocol stack to support co-existence of M2M devices and H2H users.

The main objective of this research is to investigate the challenges of M2M communications in two broad types of M2M networks; *capillary* M2M and *cellular* M2M networks. The primary focus is on developing novel solutions, algorithms, and protocol enhancements for successfully enabling M2M communications. Since cognitive radio technology is very promising for M2M communications, special emphasis is on capillary M2M networks with cognitive radio based Physical layer. Besides, the focus is also on exploring new frontiers in M2M communications.

This thesis covers different aspects of M2M communications. Considering the motivation for cognitive M2M and service requirements of M2M devices, two cognitive MAC protocols have been proposed. The first protocol is *centralized* in nature and utilizes a specialized frame structure for co-existence with the primary network as well as handling different Quality-of-Service (QoS) requirements of M2M devices. The second protocol is a *distributed* cognitive MAC protocol, which is specially designed to provide high energy efficiency and reliability for M2M devices operating in challenging wireless environments. Both protocols explicitly account for the peculiarities of cognitive radio environments. The protocols have been evaluated using analytical modeling and simulation studies.

Recently IETF has standardized a specially designed routing protocol for capillary M2M networks, known as RPL (Routing for Low Power and Lossy Networks). RPL is emerging as the *de facto* routing protocol for many M2M applications including the smart grid. On the other hand, the application of cognitive radio for smart grid communication is under active investigation in the research community. Hence, it is important to investigate the applicability and adaptation of RPL in cognitive radio environments. In this regard, an enhanced RPL based routing protocol has been proposed for cognitive radio enabled smart grid networks. The enhanced protocol provides novel modifications to RPL for protecting the primary users along with meeting the utility requirements of the secondary network.

An important challenge in LTE-based cellular networks with M2M communications is the *uplink* radio resource management as available resources are shared between M2M devices and H2H users, having different and often conflicting QoS requirements. Apart from this, energy efficiency requirements become critically important. Further, the specific constraints of Single Carrier Frequency Division Multiple Access (SC-FDMA) complicate the resource allocation problem. In this respect, an energy efficient resource allocation algorithm for the uplink of LTE networks with M2M/H2H co-existence under statistical QoS guarantees has been developed, that is based on *canonical* duality theory. The proposed algorithm outperforms classical algorithms in terms of energy efficiency while satisfying the QoS requirements of M2M devices and H2H users.

A new frontier in M2M communications is the nano-M2M communications, which is envisioned to create the Internet-of-Nano-Things (IoNT). *Molecular communication* (MC) is a promising communication technique for nano-M2M communications. In literature, no model for error performance of MC exists. Therefore, an error performance model has been developed that explicitly accounts for noise and interference effects. Since relaying and network coding based solutions are gaining popularity for nano-M2M networks, the error performance of a network coded molecular nano-M2M network has been evaluated as well.

Finally, the thesis is concluded based on the overall picture of the research conducted. In addition, some directions for future work are included as well.

Table of Contents

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Project
6LoWPAN	IPv6 over Low power Wireless Personal Area Networks
ACB	Access Class Barring
AMI	Advanced Metering Infrastructure
CDF	Cumulative Distribution Function
CDT	Canonical Duality Theory
CR	Cognitive Radio
DAG	Directed Acyclic Graph
DMC	Diffusion-based Molecular Communication
eNB	Evolved Node B
ETSI	European Telecommunications Standards Institute
GPRS	General Packet Radio Service
H2H	Human-to-Human
HAN	Home Area Network
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IoT	Internet-of-Things
IoNT	Internet-of-Nano-Things
IP	Internet Protocol
ISI	Inter Symbol Interference
ISM	Industrial Scientific and Medical
IWO	Invasive Weed Optimization
LLN	Low Power and Lossy Network

LTE	Long Term Evolution
M2M	Machine-to-Machine
MAC	Medium Access Control
MC	Molecular Communication
MTC	Machine Type Communication
MTCD	Machine Type Communication Device
MTCG	Machine Type Communication Gateway
NAN	Neighborhood Area Network
OFDMA	Orthogonal Frequency Division Multiple Access
PDF	Probability Density Function
PDR	Packet Delivery Ratio
PRMA	Packet Reservation Multiple Access
QoS	Quality-of-Service
RPL	Routing for Low Power and Lossy Networks
SCADA	Supervisory Control and Data Acquisition
SC-FDMA	Single Carrier Frequency Division Multiple Access
SNR	Signal-to-Noise Ratio
TDMA	Time Division Multiple Access
TVWS	TV White Spaces
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
WAN	Wide Area Network
Wi-Fi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WPAN	Wireless Personal Area Networks

Chapter 1

Introduction

The communication industry has seen a tremendous growth over the last two decades. A plethora of technologies exist today with a single objective of providing ubiquitous connectivity between people on the planet. The next big thing in communications would be a truly connected world of not only the people but also the everyday objects. Therefore, this decade is widely predicted to see the rise of connected devices that are not mobile phones and do not require human control. In this regard, *Machine-to-Machine* refers to the technologies that enable networked devices to exchange information among each other as well as with business applications.

1.1 Machine-to-Machine Communications

Machine-to-Machine (M2M) communications is an emerging communication paradigm that provides ubiquitous connectivity between devices along with an ability to communicate autonomously requiring no human intervention. M2M communications acts as an enabling technology for the practical realization of Internet-of-Things (IoT). The IoT is envisioned as *'a global network of connected devices having identities and virtual personalities operating in smart spaces and using intelligent interfaces to communicate within social, environmental, and user contexts'* [1]. This vision of IoT represents a future where billions of everyday objects and surrounding environments will be connected and managed through a range of communication networks and cloud-based servers [2].

Market size projections show a large potential for M2M market that is expected to grow rapidly in the next few years. This is due to a number of factors including the widespread availability of wireless technologies, declining prices of M2M modules, and economic incentives. Some of the most prominent M2M application areas include security

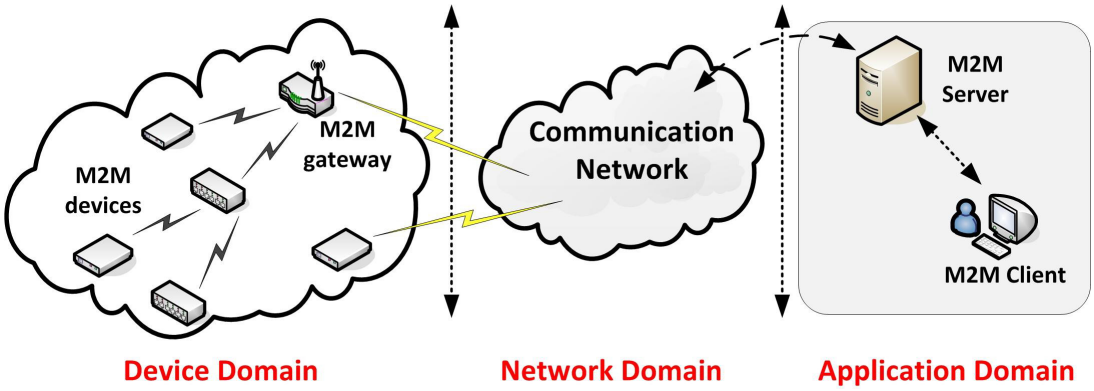


Figure 1.1: Generalized architecture of M2M communications

and public safety (surveillance systems, object/human tracking, alarms etc.), smart grids (grid control, industrial metering, demand response), vehicular telematics (fleet management, enhanced navigation, etc.), healthcare (telemedicine, remote diagnosis, etc.), manufacturing (production chain monitoring), and remote maintenance (industrial automation, vending machine control etc.) [3].

1.1.1 M2M Architecture

A generic M2M architecture consists of three distinct domains as shown in Fig. 1.1. In the *device domain*, M2M devices generally form a device area network. Devices can be connected in a single-hop or a multi-hop manner to an M2M gateway that provides interworking capabilities. M2M devices usually exchange bi-directional information with a remote server or another device. The *network domain* provides the infrastructure for realizing the communication between the device domain and the *application domain*. M2M applications contain the service middleware layer where data travels through various application services and is used by specific business processing engines. In some scenarios, M2M devices have the capabilities of autonomous interaction with the network domain.

1.1.2 Features of M2M Communications

M2M communications has unique characteristics that differentiates it from conventional Human-to-Human (H2H) communications and create new service requirements [4]. Some of

the key features and service requirements are highlighted below.

- Multitude – It is expected that the number of connected devices in an M2M ecosystem will soon exceed the sum of all those that directly interact with humans (e.g., mobile phones, tablets, etc.). An increased order of magnitude in the number of devices will generate massive transmissions, resulting in scalability issues for existing networks.
- Device Heterogeneity – There are already a large number of M2M applications and use cases due to which a large variety of devices with diverse service requirements have emerged. This creates a major challenge for interoperability as well as the generalization of M2M.
- Small Data Transmissions – Most M2M devices generate small bursty natured traffic which may have periodic patterns. Hence, the network must be able to support small amounts of data with minimal impact (e.g., signalling overhead, resource utilization, etc.).
- Energy Constrained – In most cases, M2M devices are battery operated and often deployed in areas where frequent human access and battery replacement is not easily possible. Therefore, energy efficiency for M2M communication becomes critically important in order to prolong network lifetime.
- Lossy in Nature – Reliable data delivery is equally important as energy efficiency in M2M communications because M2M devices might be deployed in challenging wireless environments where links are prone to frequent failures.
- Little or No Mobility – M2M devices are generally static, move infrequently, or move in a predefined region.
- Diversity in Quality-of-Service (QoS) – M2M communications mostly consists of low data rate applications. However, high data rate applications also exist. Similarly, latency requirements for different M2M applications vary from few milliseconds to few minutes.
- Security Vulnerabilities – A large number of connected devices in the era of IoT also means a large number of unguarded targets for hackers with increasing associated value. New security vulnerabilities arise for M2M devices and the associated networks such as physical attacks, compromise of credentials, denial-of-service attacks, user data and

identity privacy attacks, etc. Since M2M devices are running autonomously in the field, it is quite difficult to detect the attack quickly. Hence, the role of security becomes critical in M2M communications.

1.1.3 M2M Standardization

Due to rising market demands and and new business opportunities associated with M2M, different standardization activities are currently in progress [5]; mainly from 3GPP, ETSI, IEEE, and IETF. The main focus of 3GPP standardization efforts is the improvements required in the radio access network for enabling M2M communications. Similar to 3GPP, IEEE is focusing on enhancements required for 802.16 networks in order to support M2M applications. On the other hand, ETSI is covering functional architectures and service middleware layer, whereas IETF is mainly dealing with Internet connectivity solutions for M2M communications.

1.2 Scope of the Work

M2M communications creates various challenges for existing communication networks such as handling an ever increasing number of devices, managing resource constraints, providing adequate QoS, etc. Today's networks are mainly optimized for H2H communications. Hence, M2M communications must be enabled with minimum impact on H2H users. Novel solutions are needed to meet the key requirements of M2M communications in terms of scalability, energy efficiency, reliability, and diversity of QoS. Apart from this, enhancements at different layers of the protocol stack are required to support M2M communications and to optimize the networks for co-existence of M2M devices and H2H users.

M2M communications will be realized through a range of technologies and networks[1]. Generally, M2M communications is divided into two broad domains; *capillary* M2M and *cellular* M2M. In capillary M2M, connectivity among M2M devices is provided using

[1]The main focus in this thesis is on wireless M2M communications as wireless technologies will be preferred over wireline technologies due to low deployment cost.

short-range communication technologies. Wide area connectivity exists through a gateway. In cellular M2M, devices communicate autonomously with the cellular network like a typical H2H user equipment.

On the other hand, *cognitive* M2M communications[2] [7] is expected to be indispensable in the era of IoT. A large number of connected devices, as envisioned for the IoT, will create a major challenge in terms of spectrum scarcity. Through dynamic spectrum access capabilities, cognitive M2M not only improves spectrum utilization but also exploits alternate spectrum opportunities. Besides, cognitive M2M is inherently equipped to address the challenges of interference management, energy efficiency, and device heterogeneity. In addition, cognitive M2M opens new application areas for M2M communications.

The main objective of this research is to investigate different challenges of M2M communications in both capillary and cellular domains as well as to develop novel solutions, algorithms, and protocol enhancements for successfully enabling M2M communications in order to realize the vision of IoT. For capillary M2M, special emphasis is on M2M networks employing cognitive radio technology at the physical layer. Protocols for M2M networks must be designed from an application perspective. Therefore, in this thesis the main focus is on smart grid [8], which is one of the largest and most rapidly growing application area of M2M communications. Besides, the objective is to explore emerging frontiers in M2M communications as well.

1.3 Thesis Contributions

1.3.1 Key Outcomes

The contributions of this thesis cover different aspects of M2M communications. The key outcomes of this research in the form novel solutions, algorithms, and protocol enhancements are summarized below.

1. The Medium Access Control (MAC) layer plays an important role in any communication network. For successfully enabling M2M communications, designing

[2]M2M communications employing cognitive radio [6] technology for dynamic/opportunistic spectrum access.

a MAC protocol that handles massive accesses, provides high energy efficiency and reliability, supports diverse QoS requirements, and exhibits good scalability characteristics is critical. Keeping in view the promising future of cognitive M2M communications as well as specific requirements of M2M communications, a cognitive MAC protocol has been proposed for capillary M2M networks. To the best of our knowledge, no cognitive MAC protocol specifically designed for M2M communications exists in literature. The proposed protocol, which is presented in chapter 3, is a *centralized* cognitive MAC protocol and utilizes a specialized frame structure for supporting co-existence with the primary network, providing energy efficiency, optimizing the operation for periodic as well as event-driven traffic patterns, and handling a range of QoS requirements. The protocol is evaluated through analytical modeling and simulation studies. A case study is also conducted for the application of the proposed protocol in smart grid environments under the dynamics of power systems.

2. For ad-hoc M2M topologies, a *distributed* cognitive MAC protocol is proposed in chapter 4, which is specifically designed to provide high energy efficiency and reliability for M2M devices operating in challenging wireless environments. The proposed protocol uses *preamble sampling* [9] for minimizing idle listening and supporting sleep modes without synchronization. In addition, it exploits the broadcast nature of the wireless medium to provide high reliability. Apart from this, the protocol explicitly account for the peculiarities of cognitive radio environments. Analytical models and simulation studies are used for evaluating the performance of proposed protocol.
3. Due to the lossy nature of most capillary M2M networks, IETF has recently standardized a specially designed routing protocol, known as RPL (Routing for Low Power and Lossy Networks) [10], which is expected to be the standard routing protocol for majority of M2M applications. RPL is becoming increasingly popular for smart grid networks and has been under active investigation since its standardization in March 2012. On the other hand, the application of cognitive radio technology for smart grid communication is attracting a lot of attention in the research community. Hence, it is important to investigate the applicability and adaptation of RPL in

cognitive radio environments. In chapter 5, an enhanced RPL based routing protocol has been proposed for cognitive radio enabled smart grid networks. The enhanced protocol provides novel modifications to RPL, that are especially tailored for cognitive radio environments. Moreover, it adopts an opportunistic forwarding approach to provide spatial and temporal protection to primary users along with meeting the utility requirements of the secondary network. System level performance investigation is conducted to see the effectiveness of proposed protocol for smart grid network.

4. Due to ubiquitous connectivity and widespread coverage, cellular networks are expected to play an important role in future M2M networks. Cellular networks are mainly designed for H2H communications. Hence, M2M communications over cellular networks must be realized without degrading the services for H2H users. Due to higher capacity and flexible resource management, Long Term Evolution (LTE) networks are gaining more and more interest for supporting the envisioned M2M applications. The introduction of M2M communications in LTE networks creates various challenges. The most important challenge comes in radio resource management as available resources are shared among M2M devices and H2H users having different and often conflicting QoS requirements. Moreover, energy efficiency becomes critically important due to the battery operated nature of M2M devices. Since M2M devices mainly generate uplink traffic, design of uplink interface becomes challenging. In the LTE uplink, Single Carrier Frequency Division Multiple Access (SC-FDMA) is used as the multiple access scheme. The specific power and resource allocation constraints of SC-FDMA not only complicate the resource allocation problem, but also render the standard *Lagrangian* duality based resource allocation framework, as developed for Orthogonal Frequency Division Multiple Access (OFDMA), inapplicable. In chapter 6, a resource allocation algorithm for the uplink of LTE networks in M2M/H2H co-existence scenarios is proposed with special emphasis on energy efficiency and QoS requirements. The resource allocation framework is based on *canonical* duality theory [11]. Numerical results are obtained using a novel *Invasive Weed Optimization* [12] algorithm.
5. M2M communications is not just limited to the macro-scale. A new frontier in M2M communications is the *nano-M2M* communications. Recent advances in

nanotechnology enable the development of nanomachines (on the scale of one to few hundred nanometers) capable of performing simple tasks such as computing, data storage, sensing, or actuation. Hence nano-M2M communications is becoming a reality. *Nano-M2M networks* (or *Nanonetworks*) [13] i.e., network of nanomachines are envisaged to expand the capabilities of single nano-machines by allowing them to cooperate and share information. The interconnection of nanomachines with classical networks and ultimately the Internet defines a new networking paradigm, which is referred to as the Internet of Nano-Things [14]. *Molecular Communication* [15, 16] is a promising technique for nano-M2M communications. Molecular communication differs from conventional wireless communication due to its varying and dramatically high propagation delays, operational uncertainties due to random movement of molecules, extremely low reliabilities, etc. The most widely investigated molecular communication technique in literature is the *diffusion-based* molecular communication (DMC). Recently, a number of studies have investigated DMC from various aspects. Particularly, relaying and network coding based solutions have been proposed for DMC. However, no model for error performance of DMC under different kinds of impairments in general, and with network coding in particular, exists in literature. Therefore, in chapter 7, closed-form expressions for error probability have been derived for a simple nanonetwork under diffusion noise and interference effects. Additionally, the effect of different parameters on error performance is investigated along with evaluating the channel capacity of DMC.

1.3.2 List of Publications

The publications[3] related to the main contributions of this thesis are stated as follows. The chapter relevance of different publications is given in Table 1.1.

Journals

[1] **A. Aijaz** and A. H. Aghvami, "PRMA based Cognitive Machine-to-Machine Communications in Smart Grid Networks," *IEEE Transactions on Vehicular*

[3]The numbering associated with publications does not refer to the Bibliography section of this thesis.

Technology, vol. PP, no. 99, pp. 1 – 16, Sept. 2014.

[2] **A. Aijaz**, H. Su, and A. H. Aghvami, "CORPL: A Routing Protocol for Cognitive Radio Enabled AMI Networks," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1 – 9, July 2014.

[3] **A. Aijaz**, M. Tshangini, M. Reza Nakhai, X. Chu, and A. H. Aghvami, "Energy Efficient Uplink Resource Allocation in LTE Networks with M2M/H2H Co-existence under Statistical QoS Guarantees," *IEEE Transactions on Communications*, vol. 62, no. 7, pp. 2353 – 2365, July 2014.

[4] **A. Aijaz**, X. Chu, and A. H. Aghvami, "Energy Efficient Design of SC-FDMA based Uplink under QoS Constraints," *IEEE Wireless Communications Letters*, vol. 3, no. 2, pp. 149 – 152, April 2014.

[5] **A. Aijaz**, A. H. Aghvami, and M. Reza Nakhai, "On Error Performance of Network Coding in Diffusion-based Molecular Nanonetworks," *IEEE Transactions on Nanotechnology*, vol. 13, no. 5, pp. 871 – 874, Sept. 2014 (Transaction Letter).

[6] **A. Aijaz** and A. H. Aghvami, "Error Performance of Diffusion-based Molecular Communication using Pulse-based Modulation," *IEEE Transactions on NanoBioscience*, vol. PP, no. 99, pp. 1 – 6, Oct. 2014.

[7] **A. Aijaz**, S. Ping, R. Akhavan, and A. H. Aghvami, "CRB-MAC: A Receiver-based MAC Protocol for Cognitive Radio Equipped Smart Grid Sensor Networks," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4325 – 4333, Dec. 2014.

[8] **A. Aijaz** and A. H. Aghvami, "Cognitive Machine-to-Machine Communications for Internet-of-Things: A Protocol Stack Perspective," submitted to *IEEE Internet of Things Journal*, Aug. 2014.

[9] **A. Aijaz** and A. H. Aghvami, "Routing in Cognitive Radio Enabled Smart Grid AMI Networks," submitted to *IEEE Wireless Communications*, July 2014.

[10] R. Akhavan, **A. Aijaz**, S. Choobkar, and A. H. Aghvami, "On the Multi-hop Performance of Receiver Based MAC Protocol in RPL Based Low Power and Lossy Wireless Sensor Networks," accepted for *IET Wireless Sensor Systems*, March 2014.

[11] R. Akhavan, S. Choobkar, **A. Aijaz**, and A. H. Aghvami, "Adaptive Preamble Sampling Techniques for Receiver Based MAC Protocols in Lossy Wireless Sensor

Table 1.1: Publications Related to Individual Chapters

Chapter	Journals	Conferences
Chapter 2	[8]	—
Chapter 3	[1], [8]	[3], [4]
Chapter 4	[7], [8], [10], [11]	—
Chapter 5	[2], [8], [9]	[2]
Chapter 6	[3], [4]	[1], [5]
Chapter 7	[5], [6]	—

Networks," accepted for *IET Wireless Sensor Systems*, Feb. 2014.

Peer-Reviewed Conferences

[1] **A. Aijaz**, M. Reza Nakhai, and A. H. Aghvami,"Power Efficient Uplink Resource Allocation in LTE Networks under Delay QoS Constraints," accepted for *IEEE Global Communications Conference (GLOBECOM)*, Austin, TX, USA, Dec. 2014.

[2] **A. Aijaz**, H. Su, and A. H. Aghvami,"Enhancing RPL for Cognitive Radio Enabled Machine-to-Machine Networks," in proc. *IEEE Wireless Communications and Networking Conference (WCNC)*, Istanbul, Turkey, April 2014, pp. 2096 – 2101.

[3] **A. Aijaz** and A. H. Aghvami, "On the Use and Optimization of PRMA based Cognitive M2M Communications," in proc. *IEEE Global Communications Conference (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 1265 – 1271.

[4] **A. Aijaz** and A. H. Aghvami, "A PRMA based MAC Protocol for Cognitive Machine-to-Machine Communications," in proc. *IEEE International Conference on Communications (ICC)*, Budapest, Hungary, June 2013, pp. 2753 – 2758.

[5] **A. Aijaz** and A. H. Aghvami, "On Radio Resource Allocation in LTE Networks with Machine-to-Machine Communications," in proc. *IEEE Vehicular Technology Conference (VTC)*, Dresden, Germany, June 2013, pp. 1 – 5.

1.4 Outline of the Thesis

The rest of the thesis is organized as follows. Chapter 2 provides the preliminaries on the required technical background for understanding the research area addressed in this thesis.

The chapter also presents the related work on different aspects of M2M communication.

The main contributions of the thesis, which are related to three distinct areas; capillary M2M, cellular M2M, and nano-M2M, are discussed in chapters 3, 4, 5, 6, and 7. Since each contribution chapter addresses a unique research problem, concluding remarks are presented therein. Based on the overall picture of research conducted in the thesis, the main conclusions together with some directions for future work are presented in chapter 8.

Chapter 2

Preliminaries and Related Work

2.1 Machine-to-Machine Networks

Machine-to-Machine (M2M) networks can be classified into two broad domains; *capillary* and *cellular* M2M networks. In capillary M2M networks, M2M devices form a device area network wherein connectivity is provided through short range communication technologies (such as ZigBee, Wi-Fi, etc.). Wide area connectivity is provided through a gateway. Capillary M2M networks are generally characterized by huge number of low cost and low complexity devices, requirements of high energy efficiency and reliability, unplanned deployments, high packet loss ratios, use of low power link layer technologies, etc. In literature, this type of networks is also referred to as low power and lossy networks (LLNs).

In cellular M2M networks, M2M devices are equipped with embedded SIM cards and have the ability of communicating autonomously with the cellular network like a normal user equipment. Cellular M2M has unique characteristics [17] of small data transmissions, mostly mobile originated (uplink) traffic, little or no mobility of devices, service requirements of high energy efficiency, etc. Existing cellular M2M solutions are mostly GPRS based, particularly due to the low cost and convenient deployment as well as for the immediate entry of M2M business in the market. However, it is obvious that GPRS capacity is limited for supporting the envisioned M2M applications and services with hundreds of devices per cell. On the other hand, UMTS based solutions are significantly complex for low cost M2M devices. Therefore, UMTS based M2M solutions are rarely available today. The most promising candidate technology for M2M applications is 3GPP LTE as it offers higher capacity and a more flexible interface for radio resource management.

2.2 Cognitive M2M Communications

Cognitive radio technology [6] provides as a novel approach to address the spectrum scarcity and spectrum inefficiency issue in wireless networks. In cognitive radio networks, unlicensed users (secondary users) dynamically access the frequency band/channel whenever the licensed user (primary user) is absent and need to vacate the band/channel whenever the latter is detected.

There are several motivations for using cognitive radio technology in M2M communications (and hence the term cognitive M2M) [7]. Some of the main challenges of M2M networks can be successfully addressed through cognitive M2M as described below.

- Spectrum Scarcity: A fundamental challenge in M2M is the ever increasing number of M2M devices. It is expected that a multitude of connected devices (e.g., as per Ericsson 50 billion connected devices by 2020) will exist in near future. This creates a major challenge for existing communication networks in terms of spectrum congestion. With dynamic spectrum access capabilities of cognitive radio, existing spectrum can be utilized more efficiently in order to avoid the potential shortage of spectrum and to support large-scale data transmission.
- Interference: With a multitude of connected devices operating in unlicensed bands, significant interference issues will arise between self-existing and co-existing M2M networks. This will not only deteriorate the performance of M2M network, but also adversely affect the conventional Human-to-Human (H2H) services operating in the unlicensed bands. Hence, there is a need of exploring alternate spectrum opportunities such as TV White Spaces (TVWS) [18], which refer to large portions of UHF/VHF spectrum that is becoming available on geographical basis as a result of switchover from analog to digital TV. TVWS are attractive because of significant bandwidth availability (location dependent, e.g., on average $50-150$ MHz [19] in UK) and superior propagation characteristics.
- Coverage Issues: A critical issue in some M2M applications like smart grid is the huge variability in device locations. Some devices might be deployed in areas where

wireless connectivity is not always guaranteed, especially if operating in the ISM band (worldwide unlicensed band of $2.4 - 2.485$ GHz). Cognitive radio equipped M2M networks can effectively overcome this issue through dynamic spectrum access of better propagation bands such as TVWS.

- Green Requirement: A fundamental requirement in M2M communications is energy efficiency. Hence energy savings become particularly important to enhance the network lifetime. The use of cognitive radio technology has been demonstrated to be green or energy efficient as devices can adaptively adjust their transmission power levels based on operating environments [20]
- Device Heterogeneity: M2M networks are diverse in terms of applications and services which may cause diversity in network protocols and data formats. The cognitive ability is particularly suitable for M2M communication in order to deal with device and protocol heterogeneity as M2M networks will be more efficient and flexible if devices are smart enough to communicate with others freely.

Apart from addressing the technical challenges, cognitive M2M also introduces a variety of new applications [7].

2.3 Protocol Stack for Capillary M2M

A technically viable communication architecture is crucial for the realization of Internet-of-Things (IoT). The community has a consensus for the need of a standardized architecture which replaces proprietary approaches by means of a transparent end-to-end architecture. This section presents a standardized protocol stack for capillary M2M[1] communication.

The core requirements of the protocol stack for capillary M2M communications include the following [21].

- Energy Efficient Protocol Stack – Majority of M2M devices are battery operated and do not have the ability to draw power from the mains. Moreover, they are often deployed in areas where frequent human access and hence battery replacement is not always

[1]Note that the protocol stack for cellular M2M depends on the underlying cellular network.

	M2M Communications	Cognitive M2M Communications
Application	CoAP	CoAP
Transport	TCP, UDP	CR-Transport
Network	IETF RPL, IETF 6LoWPAN	CR-Routing (CORPL)
MAC	IEEE 802.15.4e, Low Power IEEE 802.11	CR-MAC, IEEE 802.15.4m
PHY	IEEE 802.15.4-2006 (ISM bands)	CR-PHY (Licensed/Unlicensed bands)

Figure 2.1: Protocol stack for (capillary) M2M communications and cognitive M2M communications

feasible. Therefore, the protocol stack must exhibit low energy consumption.

- Internet-Enabled Protocol Stack – It is of paramount importance that M2M networks are Internet Protocol (IP) enabled so that M2M devices have a universal language for communication.
- Highly Reliable Protocol Stack – Most M2M networks are inherently lossy in nature owing to operation in challenging wireless environments with frequent link failures. Hence reliability must be ensured at different layers of the protocol stack.

Fig. 2.1 shows the standardized protocol stack for capillary M2M communications. Efforts from different standardization bodies have led to the development of different protocols in order to meet the requirements of M2M networks. While a detailed discussion on different protocols is beyond the scope of this thesis, it is important to explain how these protocols meet the requirements of M2M networks.

At the Physical (PHY) layer, the most prominent standard in low power radio technology is the IEEE 802.15.4-2006 [22], which is expected to be sufficient for meeting the energy efficiency requirements of M2M devices. IEEE 802.15.4-2006 protocol is designed to operate on worldwide unlicensed frequency band of 2.4 – 2.485 GHz (ISM band). Current hardware implementations are optimized for short to medium range communication. Efforts are already underway in the CMOS[2] community to increase the range of low power sensor

[2]Complementary Metal-Oxide Semiconductor

radios [23]; a technological advancement that will bring new horizons for the IoT.

At the Medium Access Control (MAC) layer, newly developed IEEE 802.15.4e [24] protocol will be adopted. The protocol uses *time synchronized channel hopping* to combat fading and interference. It also uses a rigid slot structure with centralized or distributed scheduling to achieve high energy efficiency. Another strong candidate at the MAC layer is the low-power Wi-Fi (IEEE 802.11) [25], which promises high energy efficiency (through a power saving mode) while providing easy integration to existing infrastructure with built-in IP compatibility.

From networking perspective, IETF 6LoWPAN [26] protocol will be instrumental in connecting M2M devices to the Internet. 6LoWPAN bridges the gap between Internet and low power M2M devices by providing IPv6 networking capabilities through special encapsulation and header compression techniques that allow IPv6 packets to be sent over low power link layer technologies. Given the low power and lossy nature of M2M networks, routing issues can be very challenging. IETF has recently standardized an efficient routing protocol known as RPL (Routing for Low Power and Lossy Networks) [10], which is capable of quickly building routes, distributing routing knowledge among nodes with little overhead, and adapting topology in an efficient way. RPL is expected to be the standard routing protocol for majority of M2M applications including smart grid.

At the transport layer, conventional Transmission Control Protocol (TCP) [27] and User Datagram Protocol [28] are expected to be adopted. Last, but not the least, IETF CoAP (Constrained Application protocol) [29] will be adopted at the application layer. CoAP is based on RESTful[3] architecture [30], making it interoperable with HTTP[4] for simplified integration while meeting the specialized requirements of multicast support, low overhead, and simplicity for resource constrained M2M networks.

As mentioned earlier, cognitive M2M communications will be indispensable for the IoT. Hence, a new requirement for protocol stack arises i.e., cognitive radio enabled protocol stack. The adoption of cognitive radio at the PHY layer of M2M communications creates new challenges for upper layers. Hence, a completely new protocol stack is required with cognitive radio aware MAC, routing, and transport protocols as shown in Fig. 2.1.

[3]Representational State Transfer
[4]Hypertext Transfer Protocol

Despite active research on M2M communications over the last couple of years, cognitive M2M communications is still a vastly unexplored field with only a handful of studies. With reference to the standardization of cognitive M2M, efforts are still in infancy. Recently, IEEE has established the 802.15.4m Task Group [31] in order to specify cognitive radio aware PHY and MAC layer for cognitive M2M networks.

It should be noted that the resource constrained nature of M2M devices creates various challenges for the PHY layer design of cognitive M2M networks. Some of the main challenges include low complexity Software Defined Radio (SDR) based transceivers for energy efficient reconfigurability operations, lightweight spectrum sensing algorithms with high detection probability, and low cost dynamic spectrum access solutions that require minimum overhead.

One of the main objectives of this thesis is to develop cognitive radio aware MAC and routing protocols for cognitive M2M networks. The design of PHY layer for cognitive M2M network is beyond the scope of this thesis.

2.4 Smart Grid

The legacy electric power grid, which has lasted for years, is energy inefficient, insecure, and prone to frequent transmission failures and congestion [32]. The term *smart grid* refers to the next generation of electric grid where power distribution and management is upgraded by incorporating advanced bi-directional communications, automated control, and distributed computing capabilities for improved agility, efficiency, reliability and security [8]. It allows electricity providers, distributors, and consumers to maintain a real time awareness of operating requirements and capabilities. An integrated high performance, reliable, scalable, robust, and secure communication network is critical for the successful operation of smart grid in order to gather remote and timely information from different areas of grid equipment as well as to support different applications such as SCADA[5], smart metering, automated demand response, distribution automation, and micro grid management [33].

The smart grid communication infrastructure comprises of a multi-tier network

[5]Supervisory Control and Data Acquisition

extending across different grid operation areas including generation, transmission, distribution and consumer premises [34]. The network architecture generally comprises of three main networks: a Home Area Network (HAN) to connect domestic appliances in order to effectively manage on demand power requirements of end users, a Neighborhood Area Network[6] (NAN) that connects multiple HANs to a local access point, and a Wide Area Network (WAN) that provides a backhaul connection between utility headquarters and transmission/distribution substations.

Smart grid represents one of the largest and most rapidly growing application areas of M2M communication. It provides the foundation for realizing the Internet-of-Energy (IoE) [35]. The IoE relies on smart grid infrastructure to allow units of energy to be dispatched when and where it is required, similar to the flow of data packets in the Internet.

2.5 Nano-M2M Networks

Nanotechnology is enabling the development of devices on the scale of one to few hundred nanometers. At this scale, a nanomachine is the most basic functional unit which is defined as *a device, consisting of nano-scale components, able to perform a specific task at nano-level, such as communicating, computing, data storing, sensing and/or actuation* [13]. Nano-M2M Networks (Nanonetworks) enable nanomachines to communicate with each other and hence expand the potential applications of individual nanomachines both in terms of complexity and range of operation. The interconnection of nanomachines with other nano and macro scale devices and ultimately with the Internet enables a new networking paradigm known as the Internet of Nano-Things [14]. Although numerous applications for nanonetworks have been proposed, the biomedical applications show the best potential since the nanoscale is the natural domain of molecules, proteins and DNA sequences.

In literature, four different techniques have been proposed for nano M2M communications [13]. These are *nanomechanical*, *acoustic*, *electromagnetic*, and *molecular* communication. In nanomechanical communication, a message is transmitted through a mechanical contact between transmitter and receiver nanomachines. Acoustic

[6]In literature, NAN is also referred to as the *Advanced Metering Infrastructure* (AMI) network.

communication employs ultrasonic waves and uses acoustic energy such as pressure variation for information transfer. In electromagnetic communication, information transfer is achieved through modulation of electromagnetic waves. Molecular communication uses molecules to encode, transmit, and receive information. Among these techniques, lack of suitable communication equipment at nano-scale for nanomechanical and acoustic communication (e.g., navigation systems for the former and transducers for the latter) makes them infeasible for nanonetworks. Electromagnetic communication requires nano-scale antennas for very high frequency communication. On the other hand, molecular communication is the most promising technique for nanonetworks as nano-scale is the natural domain of molecules and hence molecular transceivers exist in nature. Moreover, molecular communication is biocompatible and has already been employed in many natural phenomenon.

2.6 Related Work

In this section, we carry out a survey of the state of the art in M2M communications. The related work on different aspects of M2M communications is summarized as follows.

2.6.1 Capillary M2M

In general, capillary M2M networks and Wireless Sensor Networks (WSNs) share some similarities. However, heterogeneity and multitude of connected devices, application oriented architecture, and a standardized protocol stack, as opposed to WSNs where a number of proprietary solutions exist, differentiate capillary M2M networks and WSNs. In literature, WSNs have been extensively investigated. Comprehensive surveys on WSNs exist in [36] and [37]. Some prominent works on capillary M2M networks are summarized as follows.

A hybrid MAC protocol for capillary M2M networks is presented in [38], which adopts a frame-based approach consisting of contention and transmission periods. To balance the trade-off between the contention and transmission period in each frame, an optimization problem is formulated, that maximizes the system throughput by finding the optimal contending probability during contention period and the optimal number of devices that

can transmit during transmission period. The protocol is shown to outperform TDMA[7] and slotted ALOHA in terms of throughput, delay, and frame utilization.

The authors in [39] present a passive synchronization based MAC (PSMAC) that synchronizes nodes in their sleep state by using interrupts generated from the proposed radio-triggered hardware. In addition to a data channel, PSMAC uses a synchronization channel which is used to activate the radio-triggered hardware as well as to disseminate the network information. PSMAC provides fast and pre-emptive slot allocation by placing the beaconing period after the contention access period. Performance comparison shows that PSMAC outperforms IEEE 802.15.4 in terms of node association time, energy efficiency, and faster data delivery at the expense of additional hardware and a dedicated channel.

The feasibility of low-power Wi-Fi for capillary M2M networks has been investigated in [25]. The IEEE 802.11 standard defines a power saving mechanism that allows the user terminals to turn off the transmitter and receiver in order to save power. In the power saving mode, the Wi-Fi access point buffers messages for user terminals that wake up periodically to listen to a beacon transmission notifying of any buffered messages. The authors conclude that the battery life of a low-power Wi-Fi enabled M2M devices depends on the operating scenario. In case of timely command messages, overall energy consumption increases due to frequent wakeups. Other operations like initialization/association, periodic data transmission, event triggered messages, and keep-alive messages for connection maintenance have smaller impact on overall power consumption, especially when high data rates are used.

RPL is currently under active investigation in the research community. Most of the work so far has focused on performance evaluation through simulation (e.g., [40–44]) and experimental (e.g., [45–47]) models. A comprehensive survey on RPL is given in [48] along with comparison with other LLN protocols.

In [49], the authors investigate the problem of running RPL over IEEE 802.15.4 MAC layer. The cluster-tree operation of 802.15.4 is modified such that nodes are able to support multiple parents in order to effectively work with the directed acyclic graph structure of RPL. The proposed scheme improves the end-to-end performance in terms of packet delivery ratio and delay.

[7]Time Division Multiple Access

The issue of lack of broadcast in RPL has been addressed in [50]. The authors present two broadcast mechanisms based on packet flooding. Simulations showed that the total retransmission overhead of broadcast messages from the root node is much lower when a node retransmits packets from all of its parents than when it retransmits only those transiting through the preferred parent.

The authors in [51] investigate sink mobility in RPL. It has been shown that using mobile sink node results in more balanced energy consumption among nodes which leads to a significant increase of network lifetime.

In [52], Dvir *et al.* identified illegitimate Version Number increase and Rank value decrease as two powerful attacks against RPL, which not only compromise the LLN traffic but also exhaust battery of nodes. They proposed a new security service for preventing these attacks.

It is expected that UDP will be widely deployed at the Transport layer for capillary M2M networks due to its low overhead compared to TCP. However, efforts are also underway to reduce the overhead of TCP for M2M networks. For example, Dunkels *et al.* [53] have presented a lightweight version of TCP implementation called Distributed TCP Caching (DTC) that uses segment caching and local retransmissions in cooperation with the link layer resulting in less control overhead.

In [54], the authors present a framework for *Lithe*, which is a lightweight implementation for secure CoAP (CoAPs). Lithe uses 6LoWPAN header compression for reducing the overhead of DTLS[8], which is the standard protocol to enable CoAPs. It has been shown that Lithe outperforms plain CoAPs as well as CoAP in terms of energy consumption and round trip time.

A performance evaluation of CoAP implemented for the embedded operating system TinyOS has been conducted in [55]. The implementation is deployed over 20 TelosB sensor nodes forming a multi-hop network. Further, RPL is implemented as the routing protocol. The authors investigate the trade-off of different performance metrics with two different objective functions for RPL.

There is a rich literature on cognitive MAC protocols. An excellent survey is given

[8]Datagram Transport Layer Security

in [56]. However, to the best of author's knowledge, no cognitive MAC protocol specially designed for M2M networks exists in literature.

2.6.2 Cellular M2M

In [57] two uplink scheduling schemes have been proposed for M2M communications in LTE networks. The first algorithm gives priority to channel quality, whereas maximum delay tolerance is the key consideration in the second algorithm . Performance evaluation is carried out in terms of bit error rate and number of effectively served requests. It has been shown that the number of the effectively served MTC devices can significantly increase if the exact delay constraints are taken into account, while dividing the devices into a limited number of QoS classes deteriorates system performance. Moreover, taking into account channel quality improves the error performance of MTC devices.

A predictive resource allocation scheme in LTE uplink for event based M2M applications has been presented in [58]. The scheme exploits correlation in traffic patterns of M2M devices. According to this scheme, whenever a device sends a scheduling request, the base station identifies neighboring devices which may benefit from a predictive resource allocation in lieu of waiting for those neighbors to send a scheduling request at their next scheduled opportunity. Results demonstrate that minimum uplink latency can be reduced from 6ms to 5ms and the mean uplink latency can be reduced by 50%.

In [59,60], a massive access management framework for QoS guarantees in LTE networks with M2M communications is presented. The proposed framework is based on organizing M2M devices in clusters according to QoS criteria (specified in terms of packet arrival rate and maximum tolerable jitter). The base station allocates a transmission interval to each cluster based on its packet arrival rate. During the transmission interval, each cluster is allocated a set of resource blocks. Performance evaluation investigates the effect of design parameters on Qos violation probability.

The authors in [61] extend the above mentioned framework and propose an analytical model which relates the packet arrival rate with the scheduling period and a specific QoS metric. The model is used to tune the transmission interval. Moreover, the periodic scheduler is enhanced with queue awareness.

The energy efficiency of LTE for small data transmissions has been investigated in [62]. The main conclusion is that the size of transport block with minimum resource allocation possible in LTE is too big for small data packets from M2M devices due to the aggressive use of Modulation and Coding Scheme (MCS) induced by Adaptive Modulation and Coding (AMC) which leads to low energy efficiency. The authors propose a solution in the form of finding optimal MCS and utilizing uplink power control.

While non-precoded Joint Transmission (non-precoded JT) and Partial Frequency Reuse (PFR) techniques are known to improve system level throughput, the authors in [63] evaluate these techniques in terms of latency performance for M2M traffic over LTE networks. It has been shown that PFR performs better than non-precoded JT and improves the average packet delay for the observed cell users and cell edge users for a range of transmit power increments.

Overload on uplink Random Access Channel (RACH) is an important issue in cellular M2M, that arises when a large number of MTC devices simultaneously access the channel. In this regard, Access Class Barring (ACB) [64] has emerged as the main solution for overload control and avoiding congestion in the access network. In ACB, the base station broadcasts a probability p_{ACB} (called ACB factor), which determines whether a user is temporarily barred from accessing the base station or not. However fixed values of p_{ACB} are not optimal.

In [65], the authors present an algorithm to dynamically adjust the ACB factor. The algorithm is based on an analytical model to determine the minimum time required to handle all the requests from the users. Performance evaluation shows that dynamic ACB achieves near optimal performance.

A congestion-aware admission control procedure is presented in [66]. The proposed scheme selectively rejects signaling messages from M2M devices at the radio access network following a probability that is set based on a proportional integrative derivative controller reflecting the congestion level of a relevant core network node. Performance evaluation shows improved performance compared to ACB based solutions.

The authors in [67] propose to combine the ETSI and 3GPP M2M reference architectures, yielding a cellular-centric M2M service architecture. The proposed architecture advocates the use of M2M relay as an M2M data aggregator to improve uplink

transmission efficiency. A tunnel-based aggregation scheme is also proposed which defines the rules for aggregating data units according to priority classes. Performance assessment shows significant reduction in protocol overheads for an LTE-Advanced network.

2.6.3 Nano-M2M

Nano-M2M communications is currently under active investigation in the research community. In [68], the authors develop a new physical end-to-end model for molecular communication. The model encompasses three distinct processes; emission process (transmitter), diffusion process (propagation), and reception process (receiver). The proposed model is analyzed in terms of normalized gain and delay as functions of system frequency and transmission range.

The authors in [69] provide in-depth overview of nanosensor technology and electromagnetic communication among nanosensors along with various nanoscale applications. They also provide a new network architecture for interconnection of nanosensors with existing communication networks.

The use of Terahertz (THz) band, with frequencies in the range of 0.1 – 10.0 THz, is proposed for electromagnetic communication among nanomachines [70]. A new propagation model is developed which is used to calculate the channel capacity of the THz channel. Results show that for very short transmission distance (order of tens of millimeters), THz channel supports very large bit rates (upto few Tbps). A graphene-based plasmonic antenna for electromagnetic communication in THz band has been proposed, modeled, and analyzed in [71]. The proposed antenna is based on a thin graphene nanoribbon and reassembles a nano-strip antenna.

In [72], the authors provide a mathematical framework to study a molecular nanonetwork by modeling it as communication over an additive Gaussian channel with noise following Inverse Gaussian distribution (AIGN). The authors obtain lower and upper bounds on the capacity of AIGN channel and discuss receiver design.

Communication via diffusion of molecules is an effective technique for information transfer in nanonetworks. In [73] two new modulation techniques, Concentration Shift Keying (CSK), and Molecular Shift Keying (MoSK), have been proposed for diffusion-based

molecular communication. Performance evaluation shows that MoSK is less susceptible to noise than CSK. Extending the work in [73], the authors in [74] proposed Molecular Concentration Shift Keying (MCSK), which is a combination of MoSK and CSK. It has been shown that MCSK outperforms both CSK and MoSK in terms of robustness.

In [75], Nakano *et al.* formulate a transmission rate control problem for molecular communication between bio-nanomachines and derive analytical expressions for throughput and efficiency as functions of transmission rate. The throughput and efficiency are shown to be in trade-off relationships for a wide range of transmission rates. The optimal transmission rates that maximize the throughput and efficiency are numerically evaluated. Further, two classes of feedback-based transmission rate control schemes are designed for autonomous bio-nanomachines to dynamically control their transmission rates, respectively based on negative and positive feedback from the receivers.

A new family of channel codes known as *ISI-free codes* has been introduced in [76], which improve the communication reliability of molecular communication. Compared to uncoded system, ISI-free coded system offers good performance with reasonably low complexity.

In [77], a molecular array-based communication (MARCO) scheme is introduced that utilizes the transmission order of different molecules to encode and exchange information symbols between nanomachines in nanonetworks. Numerical results show that MARCO provides significantly higher communication capacity, up to the order of 100 Kbps.

Chapter 3

PRMA-based Cognitive M2M Communications in Smart Grid Networks

3.1 Introduction

Cognitive Machine-to-Machine (M2M) communications will play an important role in realizing the vision of smart grid. In this chapter, after discussing the motivation of cognitive M2M communications in smart grid networks, we design a cognitive Medium Access Control (MAC) protocol, considering the unique features of M2M devices and smart grid communication requirements. We propose the use of *Packet Reservation Multiple Access* (PRMA), carry out its feasibility study, adapt, and significantly enhance it with modifications especially tailored for M2M environments. The proposed protocol is based on a *centralized* architecture, wherein a centralized controller maintains the cognitive operations for M2M devices. Moreover, it utilizes a special frame structure for supporting the co-existence of cognitive M2M network with the primary network. This frame structure is further optimized considering different trade-offs related to the primary network protection and the utility of the secondary M2M network. Performance evaluation is carried out through analytical modeling and simulation studies. We also present a case study for the application of proposed protocol in *Advanced Metering Infrastructure* (AMI) networks under the dynamics of power systems.

3.1.1 Cognitive M2M Communications in Smart Grid

Recently, the application of cognitive radio for smart grid communications has drawn much attention on different smart grid related platforms (e.g., see [78–83]). In the following we describe the motivation of cognitive M2M in different smart grid networks, along with identifying the key challenges. The interested reader is referred to Section 2.4 for a brief

overview on smart grid.

- Currently different wireless technologies such as Bluetooth, UWB (Ultra Wide Band), Wi-Fi, and ZigBee are potential candidates for the Home Area Network (HAN). However each technology has some limitations and shortcomings. For example, both Wi-Fi and UWB have high power consumption (in both transmission and reception modes) which makes them impractical for HANs [84]. Although Bluetooth has low power consumption, it is not scalable. Scalability can be provided using multiple piconets; however, this approach introduces increased latency. Compared to other technologies, ZigBee is the most favorable candidate for HAN due to its low power consumption and configuration flexibility. However ZigBee can experience interference issues from Wi-Fi, especially when Wi-Fi is operated at higher power and duty cycle than ZigBee [85]. The interference issue may become more severe when smart metering will become a mainstream application in homes. A critical issue in HANs is the huge variability in smart meter locations. Some meters may be installed at places such as inside garages, under the stairs or may be present in metal cages. Thus wireless connectivity cannot be guaranteed, especially if operating in the ISM band. This issue can be effectively resolved through dynamic spectrum access of better propagation bands (such as TV bands[1]) using cognitive radio technology.
- Next we consider the Neighborhood Area Network (NAN) which is often neglected in literature and assumed to be cellular by default. The purpose of the NAN is to connect different HANs in a neighborhood. Multiple smart meters communicate with a local access point (meter concentrator) which is further connected to Meter Data Management System (MDMS) that acts as a control center for storage, processing and management of meter data in order to be used by different applications. Depending upon the size of a utility, the number of smart meters in a network may vary from a few hundred to several thousand. Such large number of meters can create serious performance issues for cellular network especially on the uplink random access channel. Cognitive M2M in NAN provides an efficient solution in terms of providing additional bandwidth for effectively scheduling massive smart meter transmissions.

[1]Performance comparison of propagation characteristics of different bands in indoor environments can be carried out using the indoor propagation model given in [86].

- The smart grid is composed of different wireless sensor networks, located at various sites, which are used for a range of applications including remote monitoring of power generation networks, transmission and distribution networks, consumer facilities, as well as periodic measurement of different system parameters. Further, the application of wireless multimedia sensor networks (using video and acoustic sensors) can enhance the reliability, safety, and security of smart grid [87]. However this requires huge bandwidth and network resources and therefore, the use of cognitive radio technology becomes critically important for realizing such sensor applications. It has been shown that spectrum aware and cognitive sensor networks can enhance overall network performance [88].
- Last, but not the least, the application of cognitive radio in smart grid networks can also alleviate the burden of purchasing licensed spectrum for utility providers.

3.1.2 Heterogeneous Communication Requirements

The requirements of smart grid communication are mostly application driven and often heterogeneous in nature. However, primarily such communication depends on two important requirements: communication latency and large volume of messages [84]. Latency is important for making near real time decisions. It is emphasized here that the latency requirements among different applications vary and may range from few milliseconds to the order of few seconds [33]. The large volume of messages arise from the large number of connected devices and may result in congestion at different nodes in the smart grid. Apart from this energy efficiency requirements for various devices as well as sensor networks become particularly important due to their battery operated nature and possible deployment in areas where frequent human access is not always feasible. The integration of cognitive radio into smart grid infrastructure creates additional challenges. There is a need to develop dynamic spectrum access solutions with minimum control overhead and no additional hardware requirements. Moreover, joint consideration of spectrum sensing and duty cycling is required to balance the trade-off between energy efficiency and spectrum efficiency.

3.1.3 Contributions and Outline

Keeping in view the motivation for cognitive M2M and smart grid communication requirements, we develop a cognitive MAC protocol that handles massive accesses, minimizes potential causes of energy consumption for low powered devices, supports diverse QoS requirements, optimizes the operation for low data rate and periodic traffic patterns and exhibits good scalability characteristics along with efficient, low cost, low overhead dynamic spectrum access. To the best of our knowledge, no cognitive MAC protocol specifically designed for M2M communications exists in literature. In order to efficiently fulfill the communication requirements, we propose the use of PRMA, which is attractive due to a number of reasons as discussed later. We adapt PRMA for M2M environments and enhance it with novel modifications. By integrating PRMA and cognitive functionalities, a frame structure is developed for supporting the co-existence of cognitive M2M network with the primary network. We develop analytical models for various performance metrics and investigate different trade-offs. In addition to this, we present an optimization framework wherein the optimal frame structure is derived under different constraints pertaining to primary network protection as well as the utility of cognitive M2M network. We also present a case study for the application of the proposed protocol in *Advanced Metering Infrastructure* (AMI) networks and develop a scheduling algorithm for smart meters by taking into account the characteristics of power systems.

The rest of the chapter is organized as follows. Section 3.2 describes the proposed cognitive MAC protocol followed by its analytical modeling in Section 3.3. Based on the analytical models, we study different trade-offs that form the basis of optimization framework presented in Section 3.4. Section 3.5 presents numerical and simulation results as well as performance comparison with other relevant works. In Section 3.6, we present the case study for AMI networks. Finally, Section 3.7 summarizes and concludes the chapter.

3.2 Proposed Cognitive MAC Protocol

3.2.1 Motivation for PRMA

PRMA was proposed as a multiple access protocol for packet voice communications in microcellular environments. It can be viewed as a combination of slotted ALOHA, TDMA, and a reservation scheme. PRMA is suitable for speech terminals because conversational speech produces multiple packet messages during long spurts. In PRMA, time is divided into frames and each frame is further divided into time slots. The frame duration is a design variable. The basic principle of PRMA is to occupy a time slot only during speech talk spurts and release the channel during silence periods. In each frame, time slots are dynamically reserved for packets from active voice terminals. Reservation of a time slot in subsequent frames (till the end of conversation) is acquired after a contention phase which ends with a successful transmission. PRMA can also be used for data users. The unused voice slots (silence periods during conversation), as determined through speech activity detection, can be assigned to data sources. A detailed description of PRMA is beyond the scope of this chapter and the interested reader is referred to [89].

Over the last two decades, numerous publications have appeared in literature on PRMA. Some of these explore different aspects of performance through various analysis techniques (e.g., equilibrium point analysis [90] – [91], Markov Chain analysis [92], performance under slow and fast fading channels [93], evaluating the effect of handoff [94], etc.), whereas others propose different adaptations and modifications (e.g., combining PRMA and CDMA [95], applications to satellite [96] and cordless systems [97], etc.). However, the adaptability and application of PRMA for M2M communications in general as well as in smart grid networks is not investigated before.

PRMA is promising for M2M communications due to a number of factors. Firstly, due to a large number of connected devices, there is a need of schedule based MAC protocol for efficient utilization of resources. Secondly, periodic and high load traffic is most suitably handled by means of a reservation based MAC protocol. Contention based MAC protocols perform poorly under high traffic load scenarios. Most M2M devices have periodic traffic patterns; therefore, the inherent element of reservation in PRMA is directly beneficial for

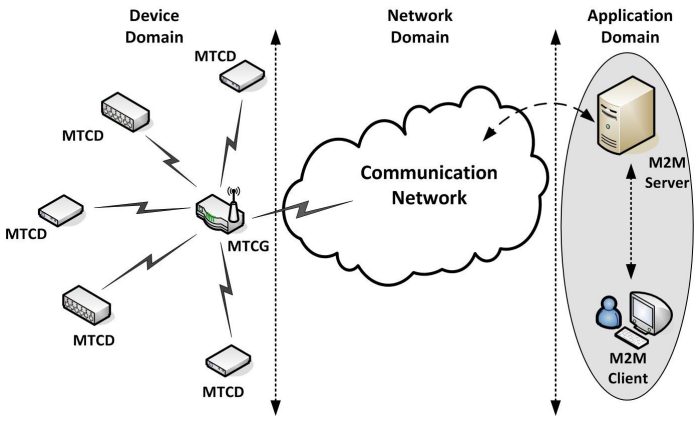


Figure 3.1: Generic M2M architecture comprising of device, network, and application domains

M2M communications. Thirdly, the TDMA based interface combined with a reservation scheme is particularly suitable from energy efficiency perspective. Energy efficiency in M2M is extremely important to prolong network lifetime. With PRMA, once a reservation has been obtained, there are no collisions, idle listening, or overhearing, which are potential causes of energy dissipation. M2M devices can stay in low power consumption mode (sleep/standby mode), waking up only at their scheduled slots and thus efficiently utilize the limited battery resources. Fourthly, due to widespread use of TDMA technology, PRMA can be economically deployed in large scale M2M networks. Last, but not the least, the TDMA based interface allows easy scheduling of network wide spectrum sensing periods for channel detection in cognitive M2M networks.

3.2.2 Network Architecture and Topology

We consider a generic M2M architecture[2] that consists of device, network, and application domains as shown in Fig. 3.1. The device domain (or M2M device area network) provides connectivity between M2M devices and the M2M gateway (which provides interworking capabilities). M2M devices communicate to a server either for periodically sending the data or in response to a request by the server. The communication network (cellular, WiMAX, satellite, etc.) provides the infrastructure for realizing this

[2]Note that this architecture reflects a wide range of smart grid communication scenarios.

communication between M2M devices and the application domain. M2M applications contain the service middleware layer where data travels through various application services and is used by specific business processing engines. In literature M2M communications is also referred to as Machine Type Communication (MTC). For the sake of brevity, henceforth M2M devices would be referred to as MTC Devices (MTCDs) and M2M gateway would be referred to as MTC Gateway (MTCG). We focus on the device domain and propose a cognitive MAC protocol for communication between MTCDs and the MTCG.

3.2.3 MAC Layer Design

It is important to mention the underlying assumptions before going into the details of the MAC protocol. We assume that the secondary network employs spectrum sensing technique for incumbent detection. Further, we assume that in our considered network topology, the MTCDs do not have any spectrum sensing capabilities. This assumption is justified considering the low cost and low complexity nature of most MTCDs. In addition, by not performing spectrum sensing, the MTCDs can keep energy consumption to a minimum. In this scenario, a low cost dynamic spectrum access solution can be realized in the form of master-slave operation i.e., another node performs spectrum sensing for MTCDs. This node is the MTCG, which acts like a centralized network controller (and therefore the proposed MAC protocol is centralized in nature) and is generally a more powerful device, free from energy constraints. The MTCG is capable of performing wide band spectrum sensing. Moreover, the MTCDs are assumed to be capable of dual band operation i.e., operation in a given licensed band and in the unlicensed band. Last, but not the least, we assume that MAC protocol is transparent to the Physical (PHY) layer i.e., the secondary network is capable of performing the required PHY layer functionalities such as the channel switching operation, which is discussed later.

In our centralized MAC protocol, the MTCG is responsible for enabling the operation of slave MTCDs. The MTCG sends an enabling signal after obtaining a vacant channel. The slave MTCDs are allowed to transmit in vacant channels. Next, we describe the frame structure as shown in Fig. 3.2. The underlying available channel (cognitive channel) is divided into a number of fixed length time slots, each able to carry a single packet. A fixed

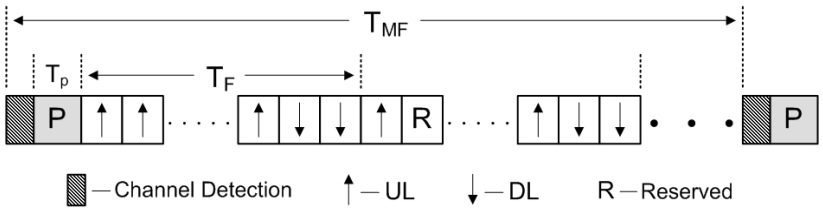


Figure 3.2: Frame structure

number of time slots are grouped together into a frame of duration T_F. In order to efficiently utilize the cognitive channel, it is preferable to use Time Division Duplex (TDD) mode of operation where both uplink (UL) and downlink (DL) are on the same channel. Since the traffic is mainly in the UL (MTCD to MTCG), the ratio of DL to UL time slots is kept small and only few time slot are assigned for any possible DL (MTCG to MTCD) communication and acknowledgements (ACKs). The DL timeslots are also used to broadcast the status (reserved or available) of UL time slots. A fixed number of frames constitute a multiframe with a duration of T_{MF}. Each multiframe starts with a channel detection period followed by a preamble or a multiframe control header that is used by the MTCG to broadcast an enabling signal carrying channel availability information. The MTCDs scan for an enabling signal. If an enabling signal is received, the MTCDs associate with the master device i.e., the MTCG.

Initially all the (UL) slots of a frame are available for contention. The contention procedure follows a slotted ALOHA scheme. Each MTCD that has data to send will contend in an available slot by transmitting a packet with some *permission probability*, p. If the packet is received correctly by the MTCG, it will send a positive ACK in the DL time slot, which also implies a reservation of the transmitted slot for the MTCD in subsequent frames. In case of a collision with another contending MTCD, the contention procedure is repeated again in another time slot as determined by a random backoff value. The backoff procedure is described later. The MTCD cannot contend again in another timeslot of the same frame unless the outcome of initial transmission is known i.e., a positive ACK from the MTCG or a collision.

To ensure fairness among the MTCDs, it is not suggested for an MTCD to keep the reservation for indefinite period of time. This is because in the steady state when all the slots

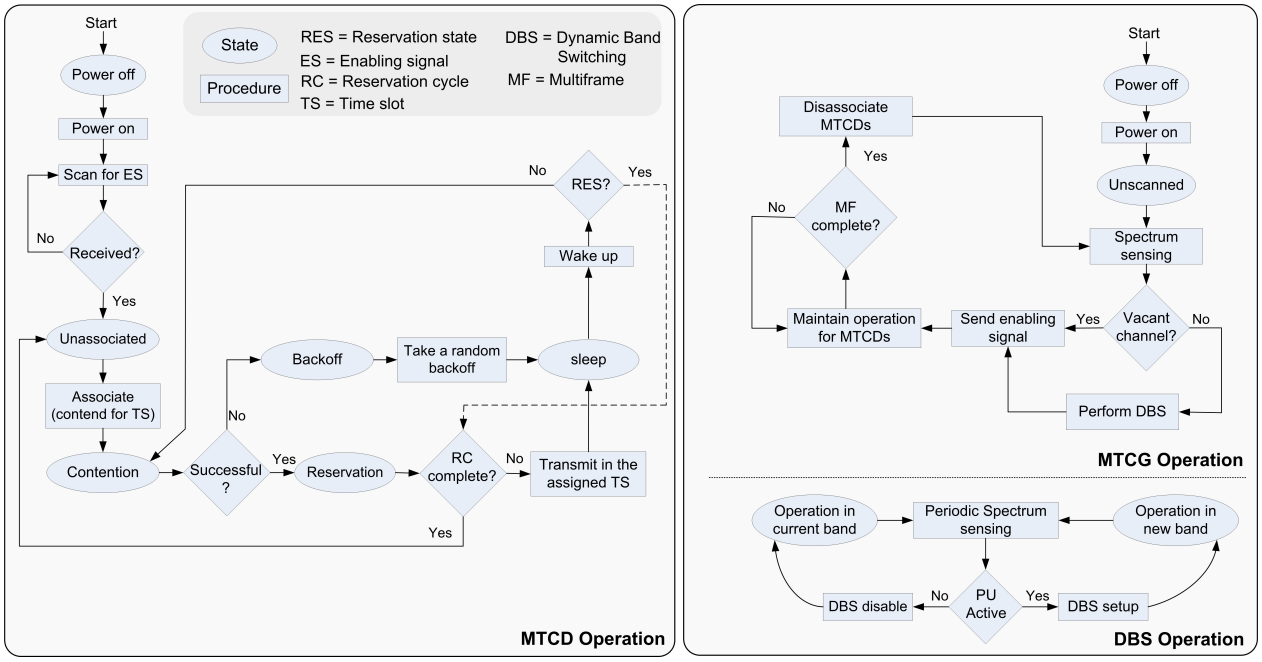


Figure 3.3: Flow charts illustrating the operation of MTCDs and MTCG in our proposed MAC layer design

of a frame are reserved, new requests would be rejected due to unavailability of resources. Secondly, certain regulatory constraints require the channel availability information to be updated periodically. Thus, there is an associated probability of change in status of the underlying available channel, when spectrum sensing is performed for the next time. Therefore, we fix the cycle of reservation for an MTCD (represented by the parameter *Reservation Cycle*) to a multiframe i.e., at the start of each multiframe, all MTCDs will contend again for reservation.

It should be noted that M2M traffic is not just limited to periodic patterns. It could also be event driven, in which case the MTCD will only transmit a single packet, not requiring a reservation. Thus, if the MTCD leaves its assigned reserved slot empty, the MTCG will interpret it as the end of transmission and broadcasts it as available in the next frame.

It is important to take into account the QoS requirements of MTCDs in resource assignment. In time slotted protocols, as in our case, this can be provided by reserving some time slots for high priority MTCDs. However this approach cannot guarantee efficient resource utilization. Keeping in view the smart grid communication requirements, we propose

a backoff procedure that incorporates the QoS requirements in the backoff taken after a collision during contention phase. This approach also improves the resource utilization. Given a total of M MTCDs, each with some QoS parameter ψ (e.g. latency/packet delay or jitter [60], which is more suitable for periodic traffic), the backoff (in terms of timeslots) taken by the m^{th} device after i^{th} collision is uniformly distributed in the interval $[1, 2, \cdots, \lambda]$, where λ is given by

$$\lambda = \frac{1}{\omega}\left\lceil 2^{\frac{\log\left(i.\frac{\psi_m}{T_s}\right)}{\delta}} - 1 \right\rceil, \tag{3.1}$$

where T_s is the slot duration, $\lceil\cdot\rceil$ denotes the ceil function, $\log(\cdot)$ denotes the *Briggsian* logarithmic function (base 10), and ω and δ are scaling factors. As it can be inferred from (3.1), devices with critical QoS requirements take a smaller backoff and hence given priority in channel access. This is also shown in the next section where a detailed analysis of backoff procedure is carried out. It should be noted that in classical PRMA and in any of its further adaptations to different scenarios, no backoff procedure exists.

The operations performed by MTCDs and MTCG in our proposed MAC layer design are illustrated by the flow charts in Fig. 3.3. In order to support co-existence with the primary network, we propose the use of *Dynamic Band Switching* (DBS) mechanism as specified in the IEEE 802.15.4m [31] standard. In DBS mechanism (illustrated in Fig. 3.3), a channel switching operation is performed whenever the primary user (PU) is detected to be active. The out of band operation can be in both licensed and unlicensed bands. However, in this paper we assume that the DBS mechanism switches transmission to unlicensed band whenever the cognitive channel needs to be vacated due to the appearance of the PU.

3.3 Analytical Modeling

In this section we develop an analytical model of the proposed MAC protocol and derive expressions for different performance indicators in cognitive M2M environments. For the ease of understanding, the analysis of cognitive operations and the dynamics of enhanced PRMA system are treated independently. However, the final expressions consider the effect of both into account. Based on analytical modeling, we investigate different trade-offs and develop

a case for optimization formulation.

3.3.1 Channel Switching Probabilities

In accordance with the DBS mechanism and the channel models used in [80] – [81], we assume that the cognitive M2M network (comprising of MTCDs and MTCG) can transmit through two different channels. The first channel lies in the licensed spectrum, referred to as the cognitive channel, which is randomly occupied by the PUs and is denoted by Ch_a. The second channel, referred to as the original channel, lies in the unlicensed spectrum and is denoted by Ch_b. The cognitive M2M network can opportunistically switch transmission to the cognitive channel. The PU activity can be modeled by a two state independent and identically distributed (i.i.d.) random process [98]. The duration of busy and idle period is assumed to be exponentially distributed with a mean of μ_{ON}^{-1} and μ_{OFF}^{-1} respectively.

Let S_b denote the state that Ch_a is busy (PU is active) with a probability $P_b = \mu_{OFF}(\mu_{OFF} + \mu_{ON})^{-1}$, and S_i denote the state that Ch_a is idle with a probability P_i, where $P_i + P_b = 1$. We assume that the MTCG employs energy detection[3] technique [99] for primary signal detection, wherein it compares the received energy (E) with a predefined threshold (σ) to decide whether Ch_a is occupied by the PU or not i.e.,

$$Sensing\ Decision = \begin{cases} S_b & \text{if} \quad E \geq \sigma \\ S_i & \text{if} \quad E < \sigma \end{cases} \tag{3.2}$$

The two principle metrics in spectrum sensing are the detection probability (P_d), and the false alarm probability (P_f). A higher P_d ensures better protection to incumbents, whereas a lower P_f ensures efficient utilization of the cognitive channel. As per [98], P_f and P_d can be expressed as

$$P_f = Pr\left\{E \geq \sigma \mid S_i\right\} = \frac{1}{2} erfc\left(\frac{1}{\sqrt{2}} \frac{\sigma - 2n}{\sqrt{4n}}\right), \tag{3.3}$$

[3]Energy detection is particularly attractive for cognitive M2M networks due to its simplicity, low signal processing cost, and minimal computational power requirements. Hence, we assume energy detection technique throughout this thesis.

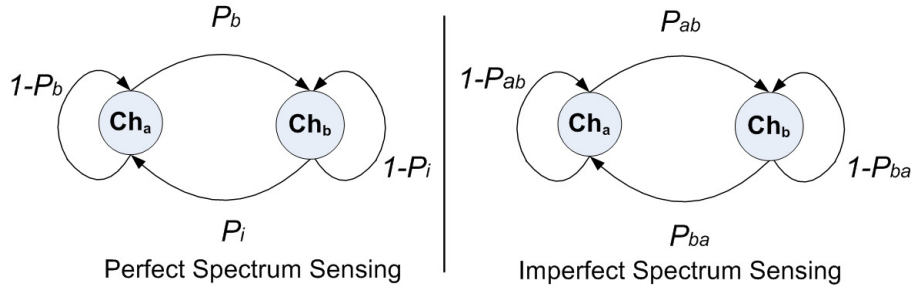


Figure 3.4: Channel switching state diagram under perfect and imperfect sensing

$$P_d = Pr\left\{E \geq \sigma \mid S_b\right\} = \frac{1}{2} erfc\left(\frac{1}{\sqrt{2}} \frac{\sigma - 2n\left(\gamma + 1\right)}{\sqrt{4n\left(2\gamma + 1\right)}}\right), \tag{3.4}$$

where $erfc(\cdot)$ is the *complementary error function*, γ is the signal-to-noise ratio (SNR) of the primary signal at the MTCG, and n is the bandwidth-time product.

We are interested in the probability of switching transmission from Ch_a to Ch_b and vice versa. Under ideal conditions, i.e., perfect spectrum sensing, these are equal to P_b and P_i respectively as shown in Fig. 3.4.

However, there can be an element of inaccuracy in spectrum sensing; therefore, the channel switching probabilities must be calculated under realistic conditions. Let P_{ab} denote the probability of switching transmission from Ch_a to Ch_b, which can be evaluated considering the following cases: (i) when S_i and a false alarm is generated i.e., $P_i P_f$; (ii) when S_b and perfectly detected i.e., $P_b P_d$. Hence,

$$P_{ab} = P_i P_f + P_b P_d \tag{3.5}$$

Similarly, let P_{ba} denote the probability of switching transmission from Ch_b to Ch_a, which can be evaluated considering the following cases: (i) when S_b and the MTCG misses to detect it i.e., $P_b(1 - P_d)$; (ii) when S_i and no false alarm is generated i.e., $P_i(1 - P_f)$. Hence,

$$P_{ba} = P_b(1 - P_d) + P_i(1 - P_f) \tag{3.6}$$

It can be easily verified that under perfect spectrum sensing i.e., $P_d = 100\%$ and $P_f =$

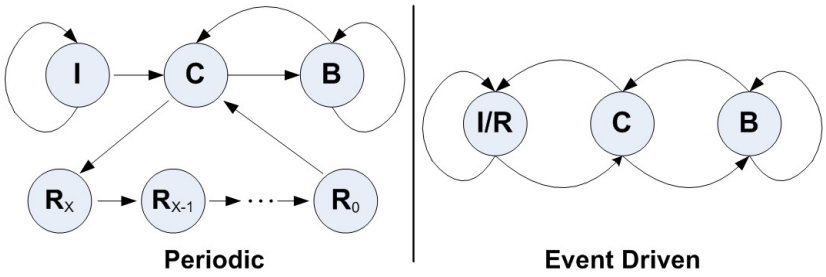


Figure 3.5: State diagrams for MTCD in our enhanced PRMA system, where I, C, B, and R correspond to Idle, Contention, Reservation, and Backoff states respectively. The subscript X with R (e.g., R_X) shows the remaining frames for keeping the Reservation state.

0%, (3.5) and (3.6) reduce to P_b and P_i respectively, which is intuitive.

After discussing the channel switching probabilities, we focus on the dynamics of our enhanced PRMA system. As shown in Fig. 3.5, each MTCD (with periodic traffic patterns) in our enhanced PRMA system, can be in Idle, Contention (CONT), Reservation (RES) or Backoff states.

3.3.2 Backoff Model

The backoff algorithm is designed especially considering the QoS requirements of the MTCDs. The objective is that under heavy load conditions (high number of reservation requests), devices with critical QoS requirements take a less aggressive backoff in events of collision and therefore access the channel with priority. This is illustrated in Fig. 3.6, where the QoS parameter is the average delay tolerance of the MTCDs. As shown, devices with lesser delay tolerance take a less aggressive backoff due to smaller backoff window.

The backoff operation for an MTCD is modeled using the state diagram shown in Fig. 3.7. Let S_j denote the j^{th} backoff state. The state transition probabilities are given as follows.

$$p_{i,0} = Pr\left\{S_{j+1} = 0 \mid S_j = i\right\} = 1 - P_c, \quad i = 0, 1, \cdots$$

$$p_{i,i+1} = Pr\left\{S_{j+1} = i+1 \mid S_j = i\right\} = P_c, \quad i = 0, 1, \cdots$$

$$p_{i,q} = 0, \qquad q \neq 0, \quad q \neq i+1,$$

where P_c denotes the probability of collision, given by $P_c = (1 - P_s)$ such that P_s represents

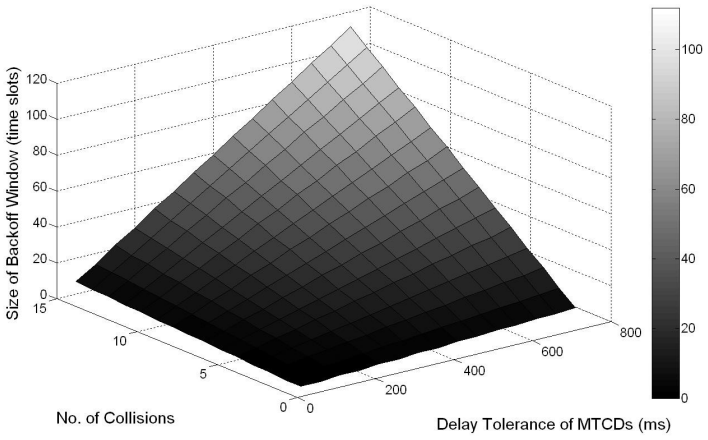


Figure 3.6: The effect of number of collisions and delay tolerance of MTCDs on the backoff interval (frame duration=100ms, no. of slots per frame=40)

the probability of successfully capturing a time slot which depends on two factors; (i) the availability of the slot (i.e., it is not reserved) and (ii) transmission from only one MTCD in that time slot. Given a total of C contending MTCDs and n_r reserved slots out of N_u uplink slots per frame, P_s is given by

$$P_s = \binom{C}{1} p\,(1-p)^{C-1} \cdot \left(1 - \frac{n_r}{N_u}\right), \tag{3.7}$$

where p represents the permission probability. It is important to mention here that a collision can also result due to PU transmission. This can happen in periodic spectrum sensing scenarios, as in our case, when the PU (which is initially detected to be inactive) returns during the transmission time of the secondary network (see the discussion on *Interference Ratio* in Section 3.3.6 for more details). Therefore, incorporating the effect of PU transmission, P_s is given by

$$P_s = \binom{C}{1} p\,(1-p)^{C-1} \cdot \left(1 - \frac{n_r}{N_u}\right) \cdot \left(1 - P_b \cdot e^{-\mu_{ON} \cdot T_s}\right), \tag{3.8}$$

where P_b denotes the probability that PU is active. Furthermore, we assume that the slot duration, T_s is small such that the PU activity does not change during this time (given by $e^{-\mu_{ON} \cdot T_s}$).

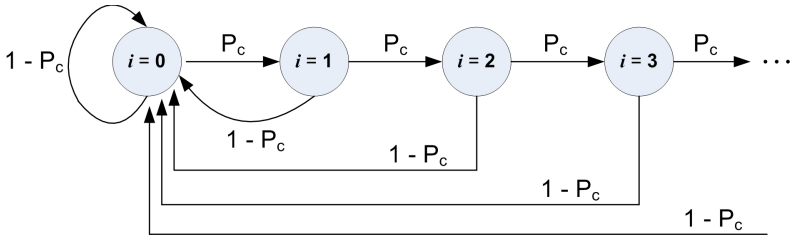


Figure 3.7: State diagram for backoff operation of MTCD

3.3.3 Average Access Delay

One of the key performance indicators is the average access delay which is defined as the average time an MTCD has to wait before obtaining a reservation. The average access delay is obtained through the analysis of the backoff operation using the backoff model described above.

In the i^{th} backoff state, the MTCD generates a random variable B_i which is uniformly distributed in the interval $[1, 2, \cdots, \lambda]$, where λ is given by (3.1). After waiting for B_i time slots, the MTCD retransmits the packet and the success or failure of transmission will determine the next state of the MTCD. Therefore, the average duration (in time slots) an MTCD stays in state i is given by $\mathbb{E}(B_i) = (1 + \lambda_i)/2$, where $\mathbb{E}(\cdot)$ denotes the expected value.

Let P_n denote the probability that MTCD will successfully transmit the packet after n collisions or alternatively after the n^{th} backoff state. Thus, P_n is given by

$$P_n = (P_c)^n P_s = (1 - P_s)^n P_s \tag{3.9}$$

Let n_t be the random variable that represents the total number of transmissions until success. Since P_n represents the Probability Mass Function (PMF) of n_t, the average number of transmissions until success can be calculated as follows.

$$N_t = \mathbb{E}(n_t) = \sum_{n=0}^{\infty} n P_n = \frac{1 - P_s}{P_s} \tag{3.10}$$

Theorem 3.3.1 *Given B_i, n_t, and N_t, the average access delay (in time slots) for an MTCD*

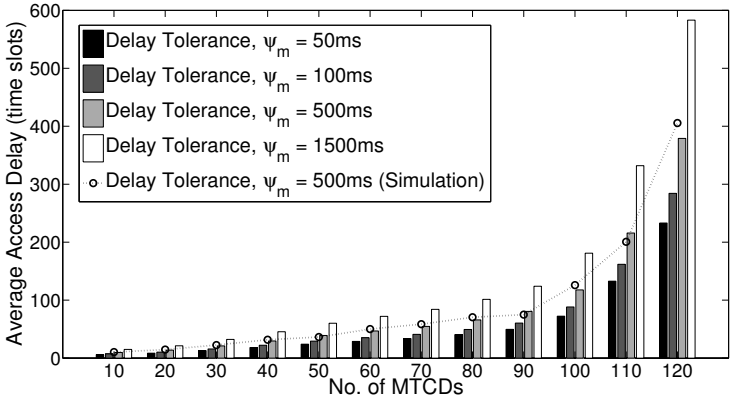


Figure 3.8: Analytical results for average access delay in time slots ($p = 0.25, N_u = 25, N = 27, T_F = 100\text{ms}, \mu_{ON} = 3, \mu_{OFF} = 1, n = U(0.1, 0.5)$). Simulation results are also shown.

is given by

$$E(w) = \mathbb{E}\left(\sum_{i=1}^{n_t} B_i\right) = \frac{1}{4}\frac{\psi_m}{T_s}\left(N_t' + N_t\right) + \frac{1}{2}N_t, \tag{3.11}$$

where $N_t' = \frac{(1-P_s)(2-P_s)}{{P_s}^2}$, ψ_m *represents the QoS indicator for the MTCD,* T_s *is the slot duration, and* N_t *is given by* (3.10).

Proof The proof is given in Appendix A.

It should be noted that (3.11) is equally valid for both periodic and event driven traffic patterns.

The analytical results for the average access delay against the number of MTCDs are shown in Fig. 3.8. The QoS requirements are given in terms of delay tolerance of the MTCDs. The average access delay increases with the number of MTCDs as the probability of collision increases, due to which an MTCD spends more time in the CONT state and consequently experiences a higher backoff time which results in delayed access to the RES state. Moreover, the MTCDs with lesser delay tolerance experience lower access delay due to less aggressive backoff window with each collision. Furthermore, for a given number of MTCDs, the access delay will increase as P_b increases due to higher probability of collision with the PU transmission. Note that the simulation results for average access delay closely follow the analytical results and hence validate the analytical modeling.

3.3.4 Throughput

We define the throughput as the ratio of the average length of the reservation interval (successful transmissions) to the cumulative average length of both reservation and contention intervals. Assuming that an MTCD encounters j successful transmissions before going back to the CONT state, the average length of the reservation interval is given by

$$\begin{aligned} E(r) &= \left[1 \cdot P_s + \sum_{j=2}^{\infty} \left(jP_r^j\right) \cdot P_{rc} \cdot \left(1 - P_b \cdot e^{-\mu_{ON} \cdot T_s}\right)\right] \\ &= P_s + \sum_{j=2}^{X-\beta} j \cdot \left(1 - P_b \cdot e^{-\mu_{ON} \cdot T_s}\right), \end{aligned} \tag{3.12}$$

where P_r and P_{rc} represent the probability of keeping the RES state and the transition from RES to CONT state respectively, and both are unity in this case. The RES state is kept only for a limited duration, the upper bound of which is determined by the *Reservation Cycle*, denoted by X. The factor β in (3.12) accounts for the average number of frames which are wasted in obtaining the reservation. This can be calculated using the average access delay derived in previous section and is given by

$$\beta = \left\lceil \frac{E(w)}{N} \right\rceil, \tag{3.13}$$

where N is the total number of slots per frame and $E(w)$ is given by (3.11). It should be noted that there is a probability of PU transmission during the reserved time slot of an MTCD, as a result of which a collision may occur during the reservation period which affects the throughput. Hence we add a factor $\left(1 - P_b \cdot e^{-\mu_{ON} \cdot T_s}\right)$ in (3.12) to account for a successful transmission as in (3.8).

The average length of contention interval represents the total number of transmissions until successfully obtaining the RES state. This is given by N_t as calculated in (3.10). Therefore, the average throughput for an MTCD is given by

$$\theta_{PRMA} = 100 \cdot \left[\frac{E(r)}{N_t + E(r)}\right] \tag{3.14}$$

It should be noted that the effect of event driven traffic is implicitly captured in throughput calculation as the first term in (3.12) is common to both periodic and event driven traffic patterns.

However, the definition of throughput is not complete unless the effect of cognitive operations is captured. By incorporating the effect of spectrum sensing, the average throughput for an MTCD is given by

$$\theta = \frac{1}{T_{MF}} \left(T_{MF} - T_d - T_p - T_{sw}\right) \cdot \theta_{PRMA}, \tag{3.15}$$

where T_{MF} is the multiframe duration, T_d is the channel detection time, T_p is the duration of preamble or multiframe control header, and T_{sw} is the average channel switching time which accounts for time overhead whenever the transmission is switched from the original channel to the cognitive channel channel (and vice versa) and is given by

$$T_{sw} = P_{ab}T_{ab} + P_{ba}T_{ba}, \tag{3.16}$$

where P_{ab} and P_{ba} are given by (3.5) and (3.6) respectively, and T_{ab} and T_{ba} are the corresponding channel switching times.

3.3.5 Duty Cycle

The duty cycle is an important performance indicator considering the energy efficiency requirements of low powered MTCDs. It is defined as the ratio of average time spent in active state (time spent in transmission and reception modes) to the total time under consideration. For the duration of one multiframe, the duty cycle for an MTCD is given by

$$DC = \frac{1}{T_{MF}} \left[T_p + N_t T_s + (X - \beta)T_s + \chi_{DL}T_s + P_{cm}T_s\right], \tag{3.17}$$

where N_t and β are respectively given by (3.10) and (3.13), and $T_{MF} = T_d + T_p + X \cdot T_F$ i.e., each multiframe consists of a channel detection time, a preamble or multiframe control header, and a total of X frames (each of duration T_F). The total transmission time for an MTCD consists of the time spent in obtaining the RES state and the successful transmission

time in the RES state. The reception time consists of listening to preamble transmissions and the DL timeslot (for monitoring the outcome of CONT state, which is abandoned once the MTCD goes to RES state). The parameter χ_{DL} in (3.17) accounts for the event count of monitoring the DL timeslot, while the MTCD is in the CONT state, and is given by

$$\chi_{DL} = \sum_{j=0}^{\infty} j\,(1 - P_c)\,P_c^j = \frac{P_c}{1 - P_c}, \tag{3.18}$$

where P_c denotes the probability of collision. In addition, the expression for duty cycle also accounts for the rare event (that occurs with a small probability P_{cm}) of any DL communication, where the MTC server sends a command message to trigger the MTCD.

3.3.6 Interference Ratio

In periodic spectrum sensing scenarios, as in our case, there is a possibility of causing harmful interference to the PU, subject to its activity. This interference is quantified in terms of *Interference Ratio* (IR), which is defined as the expected fraction of ON duration (corresponding to transmission by the PU) interrupted by the transmission of secondary users (corresponding to MTCDs in our case) [100].

The interference to PU can occur in two cases: Firstly, the PU was active; however, the MTCG misses to detect it and begins transmission. In this case the interference can occur for the duration of the secondary network transmission time (which is given by $T = T_p + X \cdot T_F$). Secondly, even if the PU was not active and the MTCG correctly detected its presence, there is a possibility that the PU activity resumes during T. Additionally, there is a possibility of one or more changes (busy and idle times) in PU activity during T. The expressions for expected interference for both cases are respectively given as follows.

$$E\,(I_b) = P_b(1 - P_d)\left[e^{-\mu_{ON}T} \cdot T + \left(1 - e^{-\mu_{ON}T}\right) P_b T\right] \tag{3.19}$$

$$E\,(I_i) = P_i(1 - P_f) \cdot \left[\left(1 - e^{-\mu_{OFF}T}\right) P_b T\right] \tag{3.20}$$

where $E\,(I_b)$ and $E\,(I_i)$ respectively denote the average interference on busy and idle states,

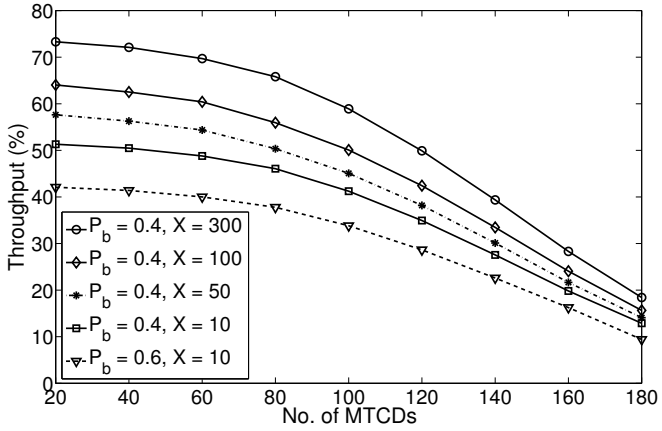


Figure 3.9: Analytical results for throughput against the device density ($p = 0.25, N_u = 25, N = 27, T_F = 100\text{ms}, T_d = 10\text{ms}, T_p = 10\text{ms}, T_{sw} = 5\text{ms}, \mu_{ON} = 3, \mu_{OFF} = 2$)

$e^{-\mu_{ON}T}$ denotes the probability that the PU activity (during the busy state) doesn't change during T, and $\left(1 - e^{-\mu_{ON}T}\right) / \left(1 - e^{-\mu_{OFF}T}\right)$ denotes the probability that the PU activity (during busy / idle states) undergoes one or more changes during T. The interference ratio is given by

$$IR = \frac{1}{TP_b} \left[E\left(I_b\right) + E\left(I_i\right)\right] \tag{3.21}$$

3.3.7 Trade-offs and the Case for Optimization Formulation

Based on the analytical models for different performance metrics derived above, we investigate some important trade-offs in the protocol design and develop a case for optimization formulation.

The results for throughput against the number of MTCDs are shown in Fig. 3.9. The average throughput per MTCD decreases with an increase in the number of MTCDs due to higher probability of collision which leads to more retransmissions, as a result of which the throughput degrades. It is particularly important to note the effect of Reservation Cycle (X) on throughput. For a given device density, as X increases the throughput increases as well. This is because the MTCD spends more time in the RES state. A lower value of X not only results in lesser time in the RES state but also causes MTCDs to contend more frequently for obtaining reservation, which results in overall degradation of throughput. Furthermore,

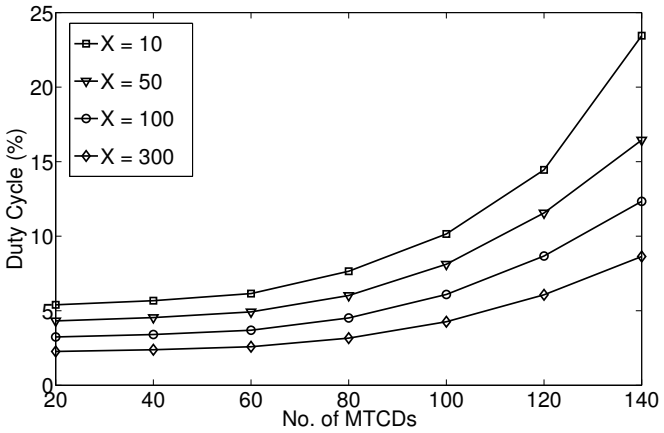


Figure 3.10: Analytical results for duty cycle ($p = 0.25, N_u = 25, N = 27, T_F = 100\text{ms}, T_d = 10\text{ms}, T_p = 10\text{ms}, T_{sw} = 5\text{ms}, P_{cm} = 0.05$)

the throughput is also dependent on the PU activity. A higher PU activity will result in throughput degradation for a given device density.

The results in Fig. 3.10 show the average duty cycle per MTCD against the number of MTCDs for different values of X. Firstly, for a given value of X, the duty cycle increases with device density as the MTCD spends more time in the CONT state and hence generate more transmissions to obtain reservation. Secondly, for a given device density, a higher value of X is desirable for a lower duty cycle as the MTCD spends more time in the RES state which means more time on average in sleep mode.

Fig. 3.11 shows the impact of reservation cycle on interference ratio for different values of busy and idle states of the PU. It should be noted that a higher value of X generates more interference to the PU as the value of X determines the transmission time of the secondary network.

Apart from the trade-offs studies above, a higher value of X may also result in unfairness among MTCDs as some MTCDs will capture the time slots for continuous transmission, denying the chance for others to transmit.

Next, we move our focus to an important trade-off in cognitive radio networks. This is the *sensing-throughput* trade-off which has attracted a lot of attention recently [101–104]. The sensing-throughput trade-off is actually composed of two trade-offs. At the physical layer it is a trade-off between the false alarm probability, P_f, and the missed detection

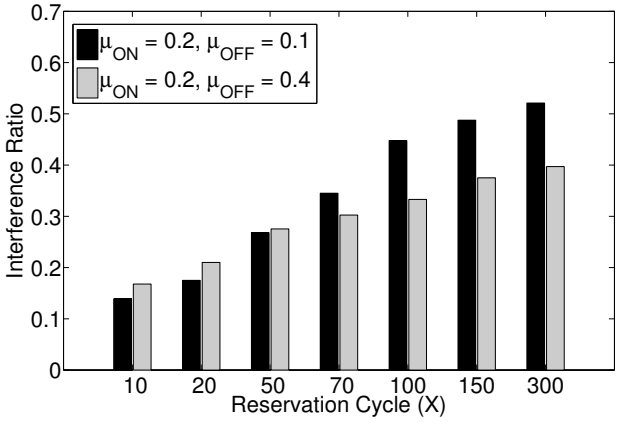


Figure 3.11: Effect of Reservation Cycle (X) on Interference Ratio ($P_d = 0.9, P_f = 0.2$)

probability, P_m, given by $P_m = 1 - P_d$, where P_d is given by (3.4). While at the MAC layer, it is the trade-off between maximizing the throughput of the secondary network and avoiding collisions with the primary network [105]. The higher the false alarm probability, the lower is the missed detection probability. However, a higher false alarm probability results in lower throughput for the secondary network as more spectrum opportunities are overlooked. On the other hand, a higher detection probability reduces collisions with the primary network and ensures better protection to incumbent users.

Keeping in view all the trade-offs discussed above, our objective is to find the optimal frame structure that not only ensures sufficient protection to incumbent users but also achieves the best performance in terms of throughput and energy efficiency for the secondary M2M network.

3.4 Optimization Framework

As stated earlier, our objective is to find the optimal frame structure that achieves the best performance for the secondary network while ensuring sufficient protection to the primary network. The optimal frame structure is found in terms of the optimal channel detection time (sensing time) and the optimal reservation cycle.

3.4.1 Optimal Channel Detection Time

Recall that the MTCG employs spectrum sensing (using energy detection) for channel detection. The objective of spectrum sensing is to achieve high spectrum re-use probability as well as high detection probability, both of which can be characterized in terms of false alarm probability (P_f) and missed detection probability (P_m). It should be noted that primary users have access priority; therefore, $P_f \geq P_m$. A higher P_f results in fewer collisions with the primary networks but at the same time the spectrum re-use probability (and hence the throughput of the secondary network) decreases as more spectrum opportunities are overlooked. The optimum trade-off is achieved with an equality relationship [106] i.e., $P_f = P_m$ (where $P_m = 1 - P_d$). Based on this the optimal channel detection time is deduced as

$$T_d^* = \frac{\sigma}{2W\left[1 + \left(\frac{\gamma}{1+\sqrt{2\gamma+1}}\right)\right]}, \tag{3.22}$$

where W is the detection bandwidth (channel bandwidth), and γ is the received SNR of the primary signal at the MTCG. The proof of (3.22) is given in Appendix B. It should be noted that the optimal channel detection time is a decreasing function of received SNR of primary signal i.e., the lower the SNR, the higher detection time is needed to achieve a certain detection probability.

3.4.2 Optimal Reservation Cycle

Based on the optimal channel detection time, the optimization framework for finding the optimal reservation cycle is given as follows.

$$\begin{aligned} (\mathcal{P}_{max}): \quad & \max_{T_d^*, X^*} \quad Y\left(T_d, T_{MF}\right) \\ s.t \quad & (a) \quad P_d \geq P_d^{'} \\ & (b) \quad IR \leq IR_{max} \\ & (c) \quad DC \leq DC_{max} \end{aligned} \tag{3.23}$$

where IR and DC are given by (3.17) and (3.21) respectively, and

$$Y\left(T_d, T_{MF}\right) = \frac{1}{T_{MF}}\left(T_{MF} - T_d - T^{'}\right) \cdot \theta_{PRMA} \cdot P_{ba}, \tag{3.24}$$

where $T^{'} = T_p + T_{sw}$, and θ_{PRMA} and P_{ba} are respectively given by (3.16) and (3.6). Since the optimal channel detection time (T_d^*) and the optimal reservation cycle (X^*) directly affect the throughput of the secondary network, the optimization of frame structure problem is converted into throughput maximization problem for the secondary network over the cognitive channel. The detection probability threshold, $P_d^{'}$, is the detection probability at SNR level as low as γ_{min}, where γ_{min} is specified by the regulator. The maximum tolerable interference by the primary network is specified by IR_{max}, where $IR_{max} \in [0, 1]$. The DC_{max} is the maximum allowed duty cycle which is an important design parameter and specified by considering the energy efficiency and hence the lifetime requirements of the network.

Next, we focus on finding the optimal reservation cycle subject to different constraints. First, we consider the detection constraint only (i.e., constraint (a) in (3.23)). It can be inferred that the throughput of the secondary network reaches its maximum when $P_d = P_d^{'}$ i.e., when the collisions with the primary network are reduced to a minimum by reliably detecting the primary signal level.

Theorem 3.4.1 *Given T_d^*, $P_d^{'}$, γ, and T_F, the optimal reservation cycle (X_d^*) that maximizes the throughput of the secondary network subject to detection constraint only is given by*

$$X_d^* = \frac{2}{\gamma T_F} exp\left(\rho + \sqrt{\frac{WT_d^*}{2}}\gamma\right)^2 \sqrt{\frac{2\pi T_d^*}{W}}\left(P_d^{'} + \alpha\right) + \frac{T_{sw}}{T_F}, \tag{3.25}$$

where $\rho = erfc^{-1}(2P_d^{'})\sqrt{2\gamma + 1}$, and $\alpha = P_b(1 - P_d^{'})P_i^{-1}$.

Proof The proof is given in Appendix C.

It should be noted that X_d^* is an increasing function of $P_d^{'}$ i.e., the more reliably the primary signal is detected, the higher is the reservation cycle and hence the throughput of the secondary network.

Next we find the optimal reservation cycle that maximizes the throughput by simultaneously considering both detection and interference constraints i.e., (a) and (b) in

(3.23). Using (3.19), (3.20), and (3.21), the interference ratio is given by

$$IR = P_m \left(P_b + e^{-\mu_{ON} T} P_i\right) + P_i \left(1 - P_f\right) \left(1 - e^{-\mu_{OFF} T}\right) \tag{3.26}$$

It has been proposed in [100] to replace the exponents μ_{ON} and μ_{OFF} in (3.26) with $\mu = \max(\mu_{ON}, \mu_{OFF})$, in order to mitigate the approximation error in the interference model which arises under certain conditions. Therefore, (3.26) can be expressed as

$$IR = P_m P_b + P_i \left(1 - P_f\right) + e^{-\mu T} \left(P_f - P_d\right) \tag{3.27}$$

We are interested in the upper and lower bounds of IR. Intuitively, for upper bound, the average busy duration of the PU equals the transmission time of the secondary network (i.e., $\mu_{ON}^{-1} = T$) and the average idle duration equals zero (i.e., $\mu_{OFF}^{-1} = 0$). Therefore, the upper bound of IR is given by $I_2 = P_m P_b + P_i(1 - P_f)$. Similarly, the lower bound (which results when the transmission time of secondary network equals zero i.e., $T = 0$) is given by $I_1 = P_m = 1 - P_d$.

Intuitively, subject to both detection and interference constraints, the optimal reservation cycle should be lower than the one obtained in (3.25), subject to detection constraint only. Therefore, one approach is to find the optimal reservation cycle from (3.27) subject to interference constraint and use the minimum of this value and the one obtained in (3.25). While this approach guarantees satisfying the interference constraint, it cannot ensure throughput maximization. In order to guarantee throughput maximization as well, we find the optimal reservation cycle subject to both detection and interference constraints as follows.

Theorem 3.4.2 *Given $P_d^{'}$, $IR_{max} \in [I_1, I_2]$, and T_F, the optimal reservation cycle (X^*) that maximizes the throughput of the secondary network subject to both detection and interference constraints is given by*

$$X^* = \frac{1}{T_F} \left[T_{sw} + \frac{2}{\gamma} exp\left(\rho^* + \sqrt{\frac{W T_d^*}{2}} \gamma \right)^2 \times \sqrt{\frac{2\pi T_d^*}{W}} \left\{ \left(\frac{1 - B(IR_{max} + P_b)}{2 + (2P_b - 1)B} \right) + \alpha^* \right\} \right], \tag{3.28}$$

where $B = P_i^{-1} exp\left[\mu(X_i^* \cdot T_F + T_p)\right]$, *such that* X_i^* *denotes the optimal reservation cycle subject to interference constraint only, given by*

$$X_i^* = \frac{1}{\mu T_F}\left\{\ln P_i - \Delta - \mu T_p + \ln(1 - 2P_d^{'})\right\}, \tag{3.29}$$

where $\Delta = \ln\left(IR_{max} + 2P_b P_d^{'} - (P_b + P_d^{'})\right)$, *and* α^* *and* ρ^* *are given as follows.*

$$\alpha^* = \frac{P_b}{P_i}\left[\frac{1 + B(IR_{max} + 3P_b - 1)}{2 + (2P_b - 1)B}\right] \tag{3.30}$$

$$\rho^* = \sqrt{2\gamma + 1} \cdot erfc^{-1}\left[\frac{2 - 2B(IR_{max} + P_b)}{2 + (2P_b - 1)B}\right] \tag{3.31}$$

Proof The proof is given in Appendix D.

It should be noted that X^* is a function of both IR_{max} and $P_d^{'}$. Lastly, we find the optimal reservation cycle by simultaneously considering detection, interference, and duty cycle constraints. Before that, a deeper insight into duty cycle is required. It should be noted that duty cycle is purely a characteristic of the MAC layer. In addition, from its formulation in (3.17), it can be easily inferred that an optimal value of reservation cycle in terms of sensing parameters cannot be obtained and thus studying its trade-off together with the interference and detection constraints (in terms of sensing parameters) is not possible. The duty cycle is actually dependent on two factors; one characterizing the CONT state (N_t) and the other characterizing the RES state (X); the latter being the dominant factor. Furthermore, from duty cycle perspective (or energy efficiency perspective) alone, X should be as high as possible. Therefore, we are interested in the maximum value of X that keeps the duty cycle below DC_{max} which is a design parameter. However, together with satisfying the detection and interference constraints, the best trade-off is achieved at $X^{'} = X^*$, where X^* is given by (3.28). Therefore, we obtain $DC^{'} = DC\mid_{X=X'}$ by fixing the factor characterizing the RES state and investigate its implications on the other factor characterizing the CONT state and DC_{max}, by evaluating the probability of violating the duty cycle constraint.

Theorem 3.4.3 *Given* $X^{'}$ *(optimal reservation cycle subject to detection and interference*

constraints), and DC_{max}, the probability of violating the duty cycle constraints is given by

$$P_v = Pr\left\{DC^{'} > DC_{max}\right\} = 1 - \frac{P_s\left[1-(1-P_s)^{\mathcal{Z}}\right]}{1-P_s}, \tag{3.32}$$

where $\mathcal{Z} = K^{'} - X^{'} + \beta - \chi_{DL} + 1$, $K^{'} = T_s^{-1}\left[(DC_{max} \cdot T_{MF}^{'}) - T_p\right]$, $T_{MF}^{'} = T_d^* + T_p + X^{'}T_F$, *and* $X^{'} = X^*$ *such that* X^* *is given by* (3.28).

Proof The proof is given in Appendix E.

3.5 Numerical and Simulation Results

3.5.1 Performance Evaluation of the Proposed Protocol

The primary objective of the optimization framework is to find the optimal reservation cycle and investigate its trade-offs and implications on different performance metrics. Therefore, we avoid the results for channel detection time against different sensing parameters such as detection probability, received SNR of the primary signal, channel bandwidth, etc. Interested readers are referred to [101] – [106]. However, it is important to mention that the optimal channel detection time is a decreasing function of primary SNR i.e., the lower the SNR, the higher detection time is needed to achieve a certain detection probability.

Fig. 3.12 shows the product of optimal reservation cycle (X^*) and the frame duration (T_F) against the threshold detection probability ($P_d^{'}$), subject to different constraints. Note that for a given $P_d^{'}$, the lower the sensed SNR (γ) of primary signal, the higher the product $X^* \cdot T_F$. Moreover, the product $X^* \cdot T_F$ is also an increasing function of $P_d^{'}$. Also note that compared to detection constraint only, $X^* \cdot T_F$ reduces subject to both interference and detection constraints. The implications of these results are further discussed as follows.

Fig. 3.13 shows the optimized throughput subject to both interference and detection constraints. For a given $P_d^{'}$, the lower the sensed SNR (γ) of primary signal, the higher is the throughput due to a higher reservation cycle. The throughput reduces for a lower $P_d^{'}$, because the probability of collision with the primary network increases which results in a lower reservation cycle (see Fig. 3.11 for the effect of reservation cycle on throughput). With

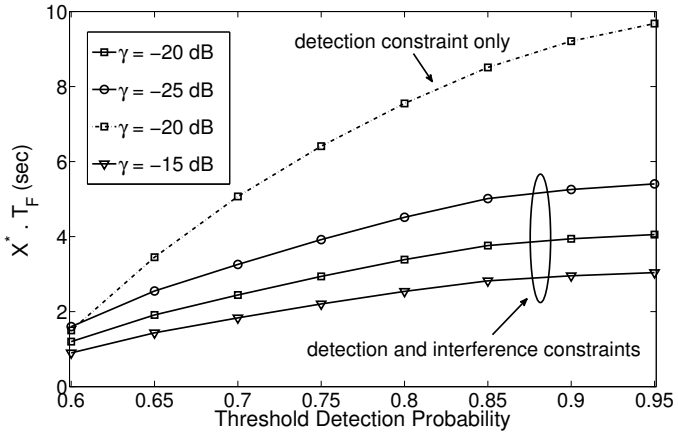


Figure 3.12: The product of optimal reservation cycle and frame duration against threshold detection probability ($W = 200\text{kHz}, T_p = 10\text{ms}, T_{sw} = 5\text{ms}, \mu_{ON} = 2, \mu_{OFF} = 3, IR_{max} = 0.5(I_1 + I_2)$)

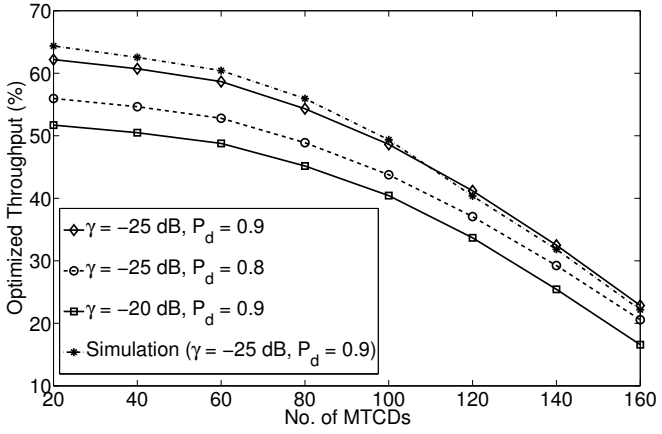


Figure 3.13: Optimized throughput subject to detection and interference constraints, against the device density ($p = 0.25, N_u = 25, N = 27, T_F = 100\text{ms}, T_d = T_d^*, T_p = 10\text{ms}, T_{sw} = 5\text{ms}, \mu_{ON} = 2, \mu_{OFF} = 3, W = 200\text{kHz}, IR_{max} = 0.5(I_1 + I_2)$). Simulation results and performance comparison is also given.

Table 3.1: Simulation Parameters for Proposed Cognitive MAC Protocol

Parameter	**Value**
Simulation duration	100,000 frames
Channel detection time (T_d)	T_d^*
Preamble duration (T_p)	10ms
Detection Probability ($P_d^{'}$)	0.9
Probability of false alarm (P_f)	0.1
Channel switch overhead (T_{sw})	5ms
Frame duration (T_F)	100ms
Permission probability (p)	0.25
QoS Parameter; Delay tolerance	Randomly varying from 10ms to 10s
Reservation cycle (X)	$X^* \mid_{\gamma=-25dB}$
Busy state parameter of PU (μ_{ON})	2
Idle state parameter of PU (μ_{OFF})	3
Maximum Interference Ratio (IR_{max})	$0.5(I_1 + I_2)$

respect to the device density, the throughput follows a similar pattern as before and decreases with an increase in device density due to higher probability of collision, which results in more retransmissions. We also conduct a MATLAB based simulation of the proposed protocol with simulation parameters given in Table 3.1 and the optimal reservation cycle dictated by Fig. 3.12. The packet arrival rates in case of periodic traffic are randomly varying from 10ms^{-1} to 1500ms^{-1}. The effect of event driven traffic is incorporated by randomly generated events, which are uniformly distributed in the interval of 10ms to maximum simulation duration. Note that he simulation results for throughput closely match the analytical results and hence validate the analytical modeling.

The results in Fig. 3.14 show the probability of violating the duty cycle constraint (P_v) against the probability of successfully capturing a time slot (P_s). Firstly, it should be noted that P_v is a decreasing function of P_s. A higher P_s results in lower time on average in the CONT state, which improves the duty cycle of MTCDs. Secondly, for a stricter duty cycle constraint (i.e. a lower value of DC_{max}), P_v increases for a given P_s (and hence a given device density). This is because satisfying the duty cycle constraint together with satisfying the interference and detection constraints for throughput maximization becomes challenging. Lastly, for a given DC_{max} and P_s, P_v increases as $P_d^{'}$ reduces due to the fact that the value of

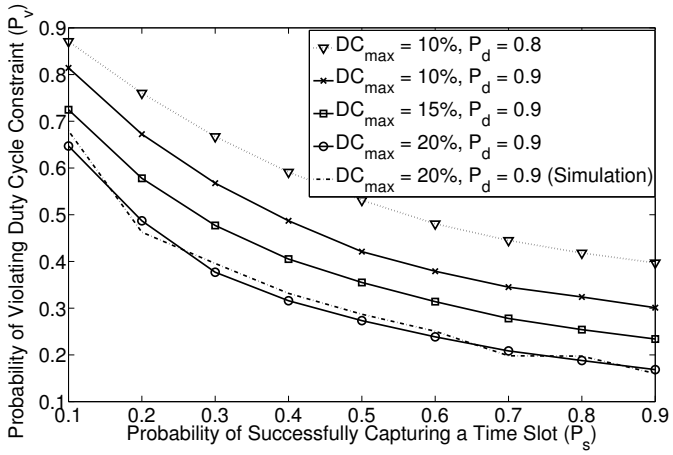


Figure 3.14: Probability of violating duty cycle constraint against probability of successfully capturing a time slot ($N = 25, T_F = 100\text{ms}, T_d = T_d^*, T_p = 10\text{ms}, T_{sw} = 5\text{ms}, \mu_{ON} = 2, \mu_{OFF} = 3, W = 200\text{kHz}, X^{'} = X^* \mid_{\gamma=-25dB}, IR_{max} = 0.5(I_1 + I_2)$). Simulation results and performance comparison is also given.

reservation cycle is reduced and hence more devices are likely to operate above DC_{max}. We also note that the simulation results closely follow the analytical results derived in Theorem 3.4.3, and hence validate the analytical model of duty cycle, given by (3.17).

3.5.2 Performance Comparison

To the best of our knowledge, no cognitive MAC protocol specifically designed for M2M communications exists in literature. Therefore, we compare the performance of different cognitive MAC protocols in M2M environments. It should be noted that our protocol is actually a *centralized* cognitive MAC protocol. Based on the classification of cognitive MAC protocols in [56], the most relevant protocol for comparison is the CSMA (Carrier Sense Multiple Access) based cognitive MAC protocol [107], which is centralized as well. However, for the sake of completeness, we also compare the performance with a *distributed* MAC protocol for cognitive radio ad-hoc networks (CRAHNs). Again, based on the classification in [56], we adopt C-MAC (Cognitive MAC) [108] which is somewhat similar to the proposed protocol due to its time slotted nature. We briefly explain the two protocols as follows.

In CSMA based cognitive MAC, the secondary network follows classical CSMA such that the nodes undertake carrier sensing for a duration of τ_s before sending a request to send

(RTS) packet to the base station (similar to MTCG). The base station may reply with clear to send (CTS) packet if the PU is detected to be absent. The C-MAC is a multi-channel cognitive MAC protocols for CRAHNs. In C-MAC, each channel is structured in the form of recurring superframes with non-overlapping Beacon Period (BP) and a Data Transfer Period (DTP). Within each superframe, C-MAC employs a slotted access mechanism such that each node is required to transmit a beacon during the BP. In C-MAC, one of the channels is selected as the Rendezvous Channel (RC) which is used for node coordination, PU detection (during Quiet Period (QP)), and resource reservation. The resource reservation is performed using distributed beaconing during the BP and follows the classical CSMA approach using RTS/CTS packets.

We conduct a MATLAB simulation based performance comparison. We adopt the same values for busy/idle duration of PU and same permission probability for MTCDs as given in Table 3.1. In case of CSMA based cognitive MAC, τ_s = 120ms, $\tau_{RTS} = \tau_{CTS}$ = 20ms. For C-MAC, we assume a superframe duration of 100ms (with BP = 20ms and DTP = 80ms), MAC frame size of 2.5Kbytes, and a data rate of 54Mbps per channel. Since C-MAC is distributed in nature, we consider variable number of nodes in an area of 2500m^2 where each node has a transmission range of 25m. Moreover, both the baseline protocols employ optimal channel detection time (T_d^*).

We compare the performance of different protocols in terms of throughput and duty cycle as the former is a universal metric for all cognitive MAC protocols and the latter is an important requirement of MTCDs. The results in Fig. 3.15a show the throughput for different protocols. For a given device density, our proposed protocol outperforms the CSMA based protocol due to its scheduling based nature. The pure contention nature of CSMA based cognitive MAC protocol cannot effectively handle the high number of accesses and therefore results in high collision probability. We note that C-MAC achieves the highest throughput under low device density (upto 20 MTCDs) and the throughput increases with the number of channels. However, with the increase in number of MTCDs, there is a sharp decline in throughput. This is because in C-MAC all the beacons sent by different nodes must be accommodated in the BP of a superframe, which results in low scalability. Another contributing factor to degraded throughput at high device densities is the CSMA based

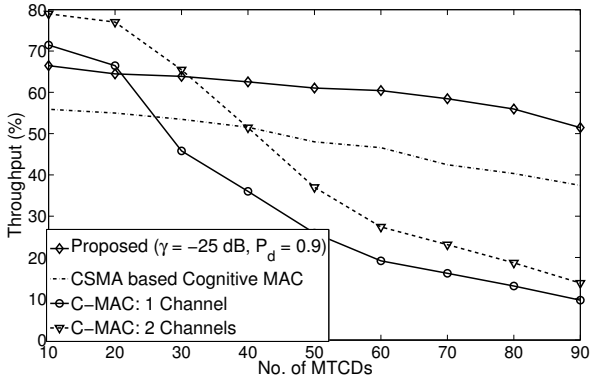
Table 3.2: Comparative Analysis of Different Cognitive MAC protocols

	Proposed MAC	**CSMA based MAC**	**C-MAC**
Architecture	Centralized	Centralized	Distributed
Synchronization Overhead	Yes	No	Yes
Scalability	High	Low	Low
Energy Efficiency	High	Low	Low
Cost of Spectrum Sensing	Low	High	High

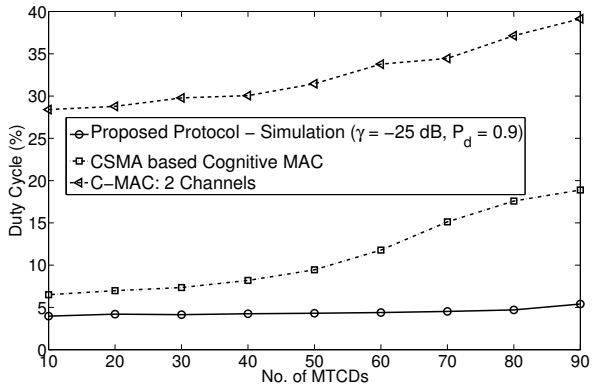
channel reservation procedure. The results in Fig. 3.15b show the duty cycle for different protocols. As seen by the results, the proposed protocol outperforms both CSMA based MAC and C-MAC due to its reservation based nature. Compared to the proposed protocol, the CSMA based cognitive MAC protocol results in higher duty cycle due to frequent channel sensing. Additionally, the collision probability is much higher when the device density increases, as a result of which the number of retransmissions increase. The C-MAC results in highest duty cycle as nodes are required to transmit beacons frequently. Moreover, frequent spectrum sensing by each node, owing to its distributed nature, also results in higher duty cycle. A comparative analysis of different protocols is given in Table 3.2

3.6 Case Study for AMI Networks under Power Systems Dynamics

In this section, we investigate the application of proposed cognitive MAC protocol in smart grid Advanced Metering Infrastructure (AMI) networks. The AMI network comprises of multiple smart meters (located at customer premises) that communicate with a local meter concentrator for transmitting various types of data, which is used by different applications. Our objective is to develop a scheduling algorithm for smart meters considering the characteristics of power systems.



(a)



(b)

Figure 3.15: Simulation based performance comparison for different protocols (a) throughput comparison, (b) duty cycle comparison

3.6.1 System Model

For simplicity, we consider a single bus[4] with one generator and K power users, each equipped with a smart meter that can perfectly measure the power consumption of the corresponding power user. We assume one access point that is able to provide wireless connectivity among all the smart meters within the range of the bus. The access point has capabilities of performing spectrum sensing operation, similar to the MTCG. In order to distinguish from the normal multiframe, the multiframe structure in this case is termed as the 'power multiframe'. The preamble in this case is used to broadcast the channel availability as well as the power pricing information. Considering the dynamics of power system (which

[4]A bus in power systems refers to a common structure connecting multiple local electrical devices. Connectivity between buses is provided by transmission lines.

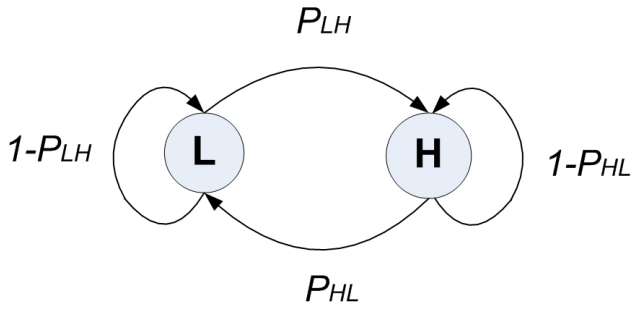


Figure 3.16: Two state Markov chain for power demand

change over the order of few hundreds of milliseconds), the power multiframe is divided into a fixed number of timeslots only. Moreover, the power multiframe itself can be scheduled periodically according to the requirements of the utility provider.

3.6.2 Model for Power Load Variation

Since the power consumer is expected to be power price aware in smart grid, the power load at each power user is dependent on the power price and the random demand for power of the user. As per [109], the reward of power user k is given by $R_k = U_k(\mathcal{P}_k, \mathcal{R}_k) - LMP \times \mathcal{P}_k$, where U_k, $\mathcal{P}_k$, and $\mathcal{R}_k$ respectively denote the utility function, power consumption, and the random power demand of user k, and LMP is the Locational Marginal Price [110].

For simplicity, we model the power demand for a user as a two state Markov chain. In practice, the power load is much more complicated. However, the Markov model simplifies the analysis and has been used widely used for modeling demand and supply in power markets (e.g., see [111–114]). An illustration of the Markov model is shown in Fig. 3.16, where the two states refer to high (H) and low (L) power demands.

3.6.3 Scheduling Algorithm

A fundamental objective in power systems is to balance the supply and demand of power. Therefore, it is important to have precise information of power load. We define the normalized generation error at the j^{th} bus as follows.

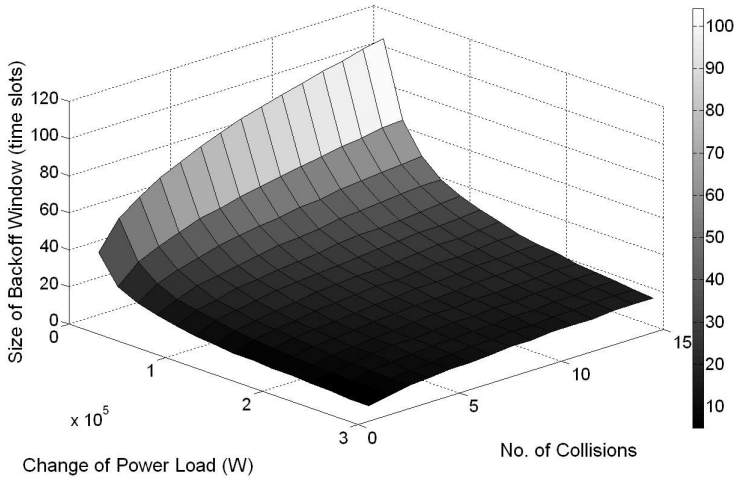


Figure 3.17: The effect of number of collisions and change of power load on the backoff interval

$$V_j = \frac{\mid G_j - G_j^* \mid}{G_j}, \tag{3.33}$$

where G_j^* denotes the power generation for the j^{th} bus based on collected load reports from smart meters and G_j denotes the optimal power generation for the j^{th} bus based on complete information of power demand (by assuming perfect communication). The objective of the scheduling algorithm is to minimize this error. Hence, it is important to schedule the smart meter having the largest expected change of power load with highest priority.

We modify the backoff procedure in our proposed protocol such that the smart meter with the largest variation in power load is more likely to win the contention procedure. The backoff (in terms of timeslots) taken by the k^{th} smart meter after i^{th} collision is uniformly distributed in the interval $[1, 2, \cdots, \lambda_p]$, where λ_p is given by

$$\lambda_p = \left\lceil \frac{\omega_p \cdot \mathcal{Z}_k}{2^{\log(i \cdot \mathcal{Z}_k)}} \right\rceil, \tag{3.34}$$

where $\log(\cdot)$ denotes the *Briggsian* logarithmic function, ω_p is a scaling constant, and $\mathcal{Z}_k$ denotes the expected change in power load of the k^{th} smart meter. As shown in Fig. 3.17, the smart meter with larger expected change of power load takes a less aggressive backoff due to smaller backoff window, and hence given priority in scheduling.

Using the two state Markov model for power load, the expected change in power load

of the k^{th} smart meter at time τ (measured in time slots) is given by

$$\mathcal{Z}_k = \mathcal{P}_k\left(L|l_{\tau'}\right)\left|\mathcal{P}_k\left(L, LMP_j\right) - \mathcal{P}_k\left(\tau - \tau'\right)\right| + \mathcal{P}_k\left(H|l_{\tau'}\right)\left|\mathcal{P}_k\left(H, LMP_j\right) - \mathcal{P}_k\left(\tau - \tau'\right)\right|, \quad (3.35)$$

where $\tau^{'}$ denotes the time elapsed since previous reporting period, $\mathcal{P}_k\left(l|l_{\tau'}\right)$, $l \in \{L, H\}$ denotes state transition probability to state l such that the state in previous reporting period was $l_{\tau'}$, $\mathcal{P}_k\left(L/H, LMP_j\right)$ is the power consumption based on current load state and locational marginal price at the j^{th} bus, and $\mathcal{P}_k\left(\tau - \tau'\right)$ is the previous power load report of the k^{th} smart meter.

3.6.4 Simulation Results

We conduct a simulation study to evaluate the performance of the proposed algorithm. For simplicity, we consider a price insensitive case. We assume that the low and high power consumption of all power users is uniformly distributed in 5KW – 15KW and 25KW – 40KW range respectively. We consider a total of 25 timeslots in the power multiframe and average the results over 1000 instances, where optimized reservation cycle is considered for multiframe duration (see Table 3.1 for multiframe parameter values).

The results in Fig. 3.18 show the normalized generation error against smart meter (power user) density per bus. We note that for a fixed state transition probability of load, a higher user density results in increased error due to higher collision probability, which leads to unsuccessful scheduling of load reports in the current power multiframe. With a similar reasoning, a higher state transition probability of load results in increased error. As expected, the proposed scheduling algorithm outperforms both CSMA based cognitive MAC and C-MAC. Moreover, a higher PU activity degrades the error performance due to higher collision probability with the PU transmission as given by (3.8).

3.7 Summary and Concluding Remarks

Cognitive M2M in smart grid networks will play an important role in realizing the vision of smart grid. Motivated by the key smart grid communication requirements of latency and large volume of messages, we propose a cognitive MAC protocol by adapting and significantly enhancing PRMA, which is attractive for M2M communications due to a number

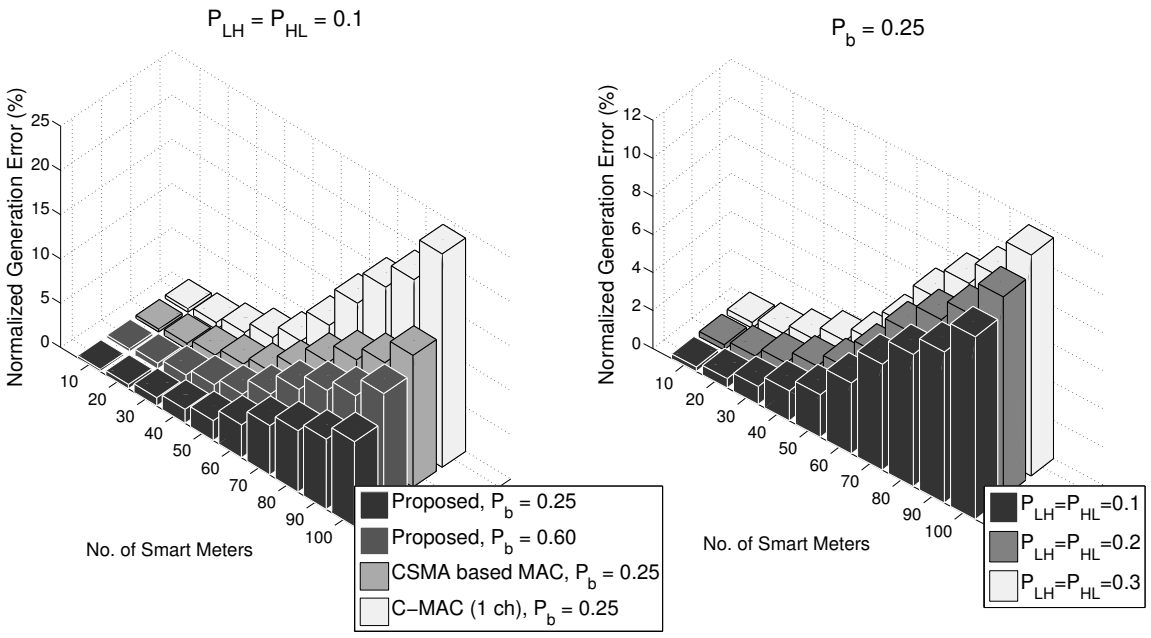


Figure 3.18: Normalized generation error against the number of smart meters for different values of PU activity and state transition probability of load

of factors. The proposed protocol is centralized in nature and provides a low cost dynamic spectrum access solution for MTCDs. A detailed investigation of different trade-offs involved in protocol design is carried out through analytical modeling. Based on the analytical models of different performance metrics, the protocol is further optimized subject to different constraints corresponding to detection and limiting the harmful interference to primary users as well as meeting the energy efficiency requirements of the MTCDs.

The protocol exhibits good scalability characteristics. Under moderate device density of $80 \sim 100$ MTCDs, throughput of up to 60% or more can be achieved depending upon the level of sensed primary signal and threshold probability of detection, while keeping the interference to incumbents to a minimum as well as satisfying the energy efficiency requirements to an appreciable extent, and thus provides a viable solution for practical M2M environments in smart grid. A case study is also presented for smart grid AMI networks under the dynamics of power systems, wherein smart meters are scheduled in order to minimize the power generation error. Simulation results show an acceptable error performance depending on user density, probability of state change in power load, and PU activity.

Chapter 4

A Receiver-based MAC Protocol for Cognitive Radio Enabled M2M Networks

4.1 Introduction

Remote and timely information gathering from different parts of grid equipment is crucial for ensuring proactive and real-time diagnosis of possible faults in smart grid. This makes cost-effective remote monitoring and control technologies vital for efficient power delivery in smart grid. In this regard, M2M networks[1] are becoming increasingly popular for the power grid. Reliable and efficient operation and management of smart grid can be accomplished with the installation of M2M devices in different parts of the grid such as distributed power plants, transmission towers and lines, substations, commercial/residential buildings, etc. Moreover, M2M video surveillance (i.e., M2M devices equipped with video and acoustic sensors) can further enhance the reliability, safety, and security of smart grid.

Wireless links in smart grid are exposed to spatio-temporally varying spectrum characteristics, which arise due to electromagnetic interference, equipment noise, dynamic topology changes, and fading [115]. Therefore, the success of smart grid operation depends on the communication capabilities of M2M devices in harsh environmental conditions that bring out great challenges for energy efficiency and reliability in M2M networks.

As discussed in Section 2.2 and 3.1.1, the use of cognitive radio (CR) technology is extremely promising for M2M communications in general and smart grid communications in particular. With dynamic spectrum access capabilities, CR enabled M2M networks can overcome the spatio-temporally varying link conditions. Moreover, access to both licensed and unlicensed spectrum bands enables the realization of envisioned multitude of connected devices without interference as well as fulfilling the additional bandwidth requirements of

[1]The term M2M networks throughout this chapter refers to ad-hoc type *capillary* M2M networks.

M2M video surveillance. Apart from this, the use of better propagation bands can effectively resolve the propagation issues due to huge variability in device locations.

Against this background, our objective in this chapter is to design a *distributed* MAC protocol for CR enabled M2M networks. In this regard, we propose CRB-MAC[2] which is a *receiver-based* MAC protocol for CR enabled M2M networks. CRB-MAC is designed with special emphasis on energy efficiency and reliability requirements of M2M networks. In order to achieve high energy efficiency, CRB-MAC employs *preamble sampling* [9] approach to tackle *idle listening* and support sleep/wakeup modes without synchronization overheads. CRB-MAC exploits the broadcast nature of wireless medium and adopts an opportunistic forwarding approach with multiple receivers as discused later in detail. This approach improves the reliability of the network along with reducing the number of retransmissions.

The rest of the chapter is organized as follows. Section 4.2 describes the framework for CRB-MAC including the system model and the protocol description. In Section 4.3, analytical models for different performance metrics are discussed. This is followed by performance evaluation in Section 4.4. Finally, Section 4.5 concludes the paper.

4.2 CRB-MAC Framework

4.2.1 CRB-MAC Overview

A key aspect of CRB-MAC is the use of *preamble sampling* for achieving high energy efficiency. In preamble sampling approach (also known as asynchronous low power listening [116]), each node selects its sleep/wakeup schedules independently of other nodes. The nodes spend most of their time in sleep mode and wake up for a short duration called *clear channel assessment* (CCA) every *checking interval* (CI) to check whether there is an ongoing transmission on the channel. To avoid deafness, the sender node transmits a long preamble with the same length as CI, followed by the data packet, to ensure that all receivers detect the preamble and obtain the data frame. By tuning CI and CCA, average duty cycles of up to 1% can be achieved without any need for scheduling or synchronization.

[2]Cognitive Receiver-Based MAC

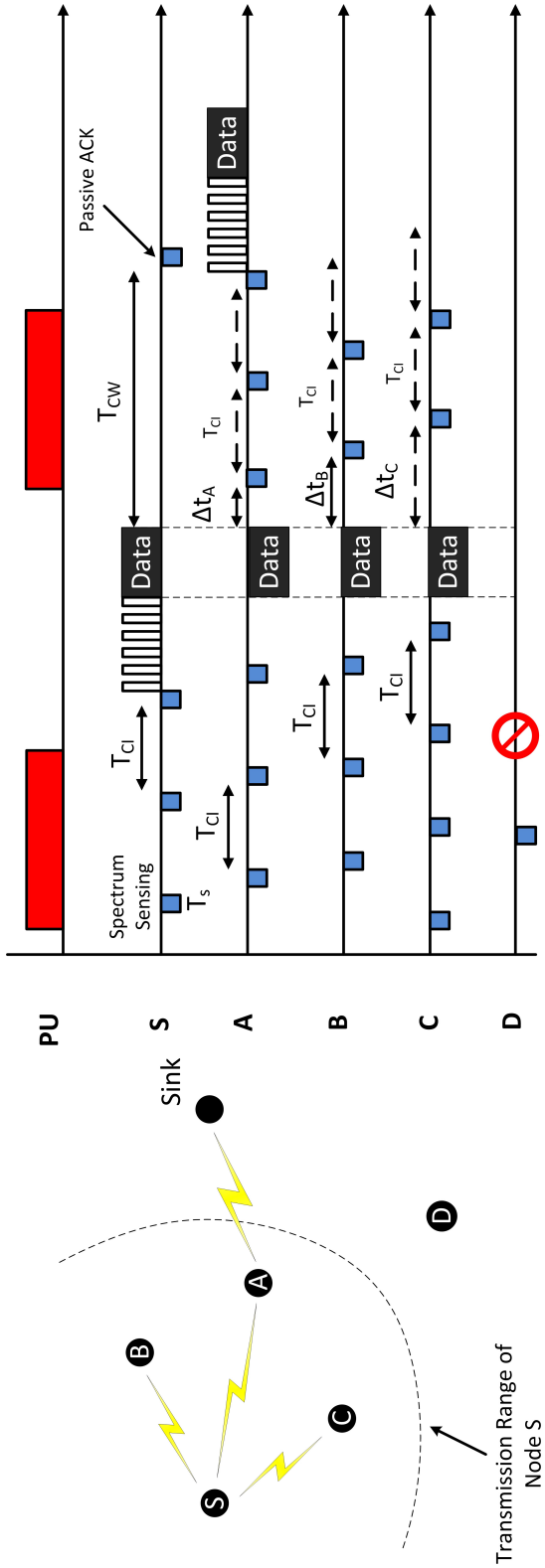


Figure 4.1: Timeline of CRB-MAC protocol with an illustrated scenario of sender and receiver nodes

CRB-MAC is inherently *receiver-based* in nature. Unlike *sender-based*[3] MAC protocol, in receiver-based MAC protocol, a sender node transmits its data without defining a particular node as a receiver. All the neighboring nodes within communication range of the sender node receive the data packet. Based on the information received from the preamble, each individual node decides if it is eligible to participate in forwarding the data. Receivers compete in an *elective* process and the winner forwards the data to the next hop towards gateway/sink.

Spectrum sensing is a key aspect of any CR environment. Nodes periodically monitor the current channel for primary user (PU) activity before using it for transmission. During this interval (sensing time) nodes are not involved in forwarding data packets, as a result of which overall network performance degrades (e.g., in terms of end-to-end throughput, latency, and packet loss ratio). CRB-MAC utilizes a mechanism to improve overall network performance under spectrum sensing state of different nodes. Further, in CRB-MAC, nodes employ optimal transmission time subject to an interference constraint, in order to ensure protection to PUs.

4.2.2 Network Model

We consider an ad-hoc network of stationary M2M devices that are CR enabled[4]. Each device (node) is equipped with a single radio transceiver that can be tuned to any channel in the licensed spectrum. We assume J stationary PU transmitters (and hence J available channels) with known locations and maximum coverage ranges. The PU (transmitter) activity model for the j^{th} channel is given by a two state independent and identically distributed (i.i.d.) random process such that the duration of busy and idle periods is exponentially distributed with a mean of $\frac{1}{\mu^j_{ON}}$ and $\frac{1}{\mu^j_{OFF}}$, respectively. Let S^j_b denote the state that the j^{th} channel is busy (PU is active) with probability $P^j_b = \frac{\mu^j_{OFF}}{\mu^j_{ON}+\mu^j_{OFF}}$, and S^j_i the state that the j^{th} channel is idle with probability P^j_i, such that $P^j_i + P^j_b = 1$. We assume

[3]In sender-based MAC (such as 1-hopMAC [117] protocol), a node that has data to send selects a receiver node from its neighbor table, includes the receiver's address in the packet header, and transmits the packet.

[4]The resource constraint nature of M2M devices necessitates the development of low cost dynamic spectrum access solutions for (ad-hoc) cognitive M2M networks. However, this is beyond the scope of this chapter.

that a node employs energy detection technique [99] for primary signal detection wherein it compares the received energy (E) with a predefined threshold (σ) to decide whether the j^{th} channel is occupied or not i.e.,

$$Sensing\ Decision = \begin{cases} S_b^j & \text{if} \quad E \geq \sigma \\ S_i^j & \text{if} \quad E < \sigma \end{cases} \tag{4.1}$$

The two principle metrics in spectrum sensing are the detection probability (P_d), and the false alarm probability (P_f). A higher detection probability ensures better protection to incumbents, whereas a lower false alarm probability ensures efficient utilization of the channel. As per [98], false alarm and detection probabilities for the j^{th} channel can be expressed as follows.

$$P_f^j = Pr\left\{E \geq \sigma \mid S_i^j\right\} = \frac{1}{2} erfc\left(\frac{1}{\sqrt{2}} \frac{\sigma - 2n_j}{\sqrt{4n_j}}\right), \tag{4.2}$$

$$P_d^j = Pr\left\{E \geq \sigma \mid S_b^j\right\} = \frac{1}{2} erfc\left(\frac{1}{\sqrt{2}} \frac{\sigma - 2n_j\left(\gamma_j + 1\right)}{\sqrt{4n_j\left(2\gamma_j + 1\right)}}\right), \tag{4.3}$$

where $erfc(\cdot)$ is the *complementary error function*, and γ_j and n_j denote the signal-to-noise ratio (SNR) of the primary signal and the bandwidth-time product for the j^{th} channel respectively.

The MAC frame structure in a CR network consists of a sensing slot (T_s) and a transmission slot (T). In periodic spectrum sensing scenarios, there is a possibility of causing harmful interference to PUs due to imperfect spectrum sensing in realistic conditions. This interference is quantified in terms of *Interference Ratio* (IR), defined as the expected fraction of ON duration of PU transmission interrupted by the transmission of secondary users and is given for the j^{th} channel as follows [100].

$$IR_j = \left(1 - P_d^j\right) P_b^j + P_i^j \left(1 - P_f^j\right) + e^{-\mu T}\left(P_f^j - P_d^j\right), \tag{4.4}$$

where $\mu = \max(\mu_{ON}^j, \mu_{OFF}^j)$. We assume that the nodes in our network employ optimal transmission time that maximizes the throughput of the secondary network subject to an

interference constraint i.e., $IR_j \leq IR^j_{max}$, where IR^j_{max} denotes the maximum tolerable interference ratio on the j^{th} channel. This transmission time for the j^{th} channel is given as follows.

$$T_j = \mu^{-1}\left[\ln P_i^j - \ln\left(P_i^j P_d^{'} + P_b^j(1-P_d^{'}) - IR^j_{max}\right) + \ln(2P_d^{'}-1)\right], \tag{4.5}$$

where $P_d^{'}$ is the detection probability threshold, defined as the detection probability at SNR level as low as γ_{min}, where γ_{min} is specified by the regulator.

4.2.3 Protocol Description

Fig. 4.1 illustrates the CRB-MAC protocol operation along with the timeline for different nodes. As shown in the figure, a node S wants to send data to the sink/gateway node by forwarding towards its first hop neighbors (within the transmission range). Firstly, it performs spectrum sensing (with duration given by T_s) to detect any PU activity. If the channel is detected as busy with PU transmission, the sender node goes to sleep mode. The sensing operation is repeated after a duration of checking interval (T_{CI}). If the PU is detected to be absent, the node starts transmitting the preamble followed by the data. The preamble consists of multiple micro-frames (each of duration T_m) and contains identification information for neighboring nodes to distinguish between PU transmission or M2M device transmission. All the nodes within the transmission range of S detect and sample few micro-frames of the preamble to extract necessary information (e.g., sequence number of the data). As shown in Fig. 4.1, only three neigboring nodes of S (i.e., nodes A, B, and C) are eligible to forward the data towards the sink node. They all wake up and receive the data transmitted by node S. If the received data packet is detected to be erroneous, it is simply discarded. The nodes receiving the data packet do not send any acknowledgement (ACK) message. However, they set a timer (Δt) before forwarding the data to the next hop. The timer is set relative to a node's distance from the sink. The distance is determined by the network layer. For example in case of RPL [10] protocol, each node is assigned a *rank* based on an *objective function*, that determines a node's virtual position with respect to the sink node. The node with the shortest timer (closest to the sink) is most likely to forward the data towards the sink. Right after the expiry of the timer, each neighboring node performs the sensing operation. If the channel is occupied by the PU, the node goes

back to sleep mode for a duration of T_{CI}. However, if an M2M device transmission is detected, each node compares the sequence number of the transmitted data with its own. If the sequence numbers match, it means that the same data is being transmitted by another node. Therefore, it discards the data packet. Otherwise, a free channel indicates that this node is the winner and can start transmitting the preamble (e.g., in Fig. 4.1, node A is the winner). The sender node S retransmits the data if none of the participating nodes in the contention window is successful to forward the data packet. The sender node can realize this by performing the sensing operation just before ending the contention window (passive ACK). The duration of contention window, T_{CW}, is set according to the transmission radius of sender nodes. In case of multiple hops, the same operation continues until the data is received by the sink.

CRB-MAC uses a technique for mitigating the performance degradation due to spectrum sensing. The key aspect of this technique is to improve the performance by reducing the spectrum sensing time. Reduction of sensing time is possible when a node is situated in region of low PU activity, and hence the number of channel changes that occur over time is small [118]. Initially the sensing time is set to a maximum value i.e., $T_s = T_s^{max}$ for a fixed missed detection probability ($P_m^j = 1 - P_d^j$). The sensing time is decreased over time (by tracking the PU activity and establishing the fact that the node is located in region of low PU activity) according to the following relation: $T_s^{new} = T_s - \varphi \cdot \Delta_s$, where Δ_s is the step size, given by $\Delta_s = 0.5 \times T_s$ and φ is a constant which is obtained from the gradient of sensing time versus the missed detection probability curve (see [118] for more details). When successive missed detection events occur, the node increases the sensing time with similar step size.

In general smart grid traffic comprises of two types: low priority monitoring data (that can be considered as best-effort) and high priority delay sensitive alarms that have an associated deadline. In order to successfully handle delay sensitive traffic, we propose an enhancement to CRB-MAC with a deadline aware forwarding process (and hence termed as deadline-aware CRB-MAC) wherein the node that provides the highest delay budget margin forwards the packet. We assume that the time before the deadline expires can be uniformly shared among the nodes in the route. Hence, the delay budget for the transmission of packet

P is given by

$$\mathfrak{D}(P) = \frac{1}{h(k)} \cdot [\Theta(P) - t_c], \tag{4.6}$$

where $\Theta(P)$ is the deadline associated with the packet P, t_c is the current time, and $h(k)$ is the hop distance between the k^{th} node and the sink node. We assume that the node is aware of the hop count to sink node through network layer information exchange. When a packet is at node k, the delay before the packet is correctly transmitted to the next hop depends on: (a) the average delay until a vacant channel is found (d_1) and (b) the average delay until the next hop correctly receives the packet (d_2). While d_1 depends on PU activity and can be estimated using (4.16), d_2 is characterized by the MAC layer and can be estimated using link metrics such as ETX (*Expected Transmission Count*[5]) [119].

In order to guarantee that the packet is forwarded to the sink node before the deadline, it is important to give priority to the forwarding node that provides the highest margin for the delay budget ($\mathfrak{D}(P) \geq d_1 + d_2$). Therefore, in deadline-aware CRB-MAC, the key difference is that the timer (Δt) is set according to the delay budget i.e., a node which provides a higher delay budget sets a lower timer and vice versa). The deadline information is embedded in the preamble transmission and therefore, available to the neighboring nodes during the checking interval. If the deadline has elapsed after the expiry of timer (Δt) and finding a vacant channel, the receiving node drops the packet.

4.3 Analytical Modeling

4.3.1 Probability of Channel Switching

We are interested in probability of switching transmission to the cognitive channel. The CR users can only use the licensed channel in the absence of PU activity. However, under realistic conditions, there can be an element of inaccuracy in spectrum sensing. Let P_{sw}^{j} denote the probability of switching transmission to the j^{th} cognitive channel which can be evaluated considering the following cases: (i) when S_b^j and the node misses to detect it; (ii)

[5]The ETX of a link between nodes k and l is given by $ETX_{kl} = 1/p_{kl}$, where p_{kl} is the probability of node l receiving a transmission from node k.

when S_i^j and no false alarm is generated. Hence P_{sw}^j is given by

$$P_{sw}^j = P_b^j \left(1 - P_d^j\right) + P_i^j \left(1 - P_f^j\right) \tag{4.7}$$

It can be easily verified that under perfect spectrum sensing conditions i.e., $P_d^j = 100\%$ and $P_f^j = 0\%$, (4.7) reduces to P_i^j, which is intuitive.

4.3.2 Energy Consumption and Retransmission Model

In CRB-MAC, the probability of failure of a single transmission on the j^{th} channel depends on the corruption in preamble or data frame and is given by

$$P_{fail}^j = P_{sw}^j \left[1 - (1-p)^{m+d}\right], \tag{4.8}$$

where m and d respectively denote the size of micro-frame and data frame in bits, p denotes the bit error probability, and P_{sw}^j is given by (4.7).

Let, $r_m = \left\lceil \frac{T_{pr}}{T_m} \right\rceil$ denote the number of micro-frames in the preamble, where T_{pr} denotes the preamble duration and T_m is the transmission time for one micro-frame. On the transmitter side, the expressions for energy drained in a single successful and failed transmission on the j^{th} channel are given by

$$\mathcal{E}\_T_{succ}^j = \mathcal{E}_{ss}^j + P_{sw}^j \left\{(1-p)^m r_m T_m + (1-p)^d T_d\right\} \mathcal{P}_t, \tag{4.9}$$

$$\mathcal{E}\_T_{fail}^j = \mathcal{E}_{ss}^j + P_{sw}^j \left\{r_m T_m + \left(1 - (1-p)^d\right) T_d\right\} \mathcal{P}_t, \tag{4.10}$$

where $\mathcal{P}_t$ denotes the power drained in the transmit mode, T_d is the duration of the data frame, and $\mathcal{E}_{ss}^j$ denotes the energy drained during spectrum sensing such that $\mathcal{E}_{ss}^j = (\tau + T_s)\mathcal{P}_s$ where $\mathcal{P}_s$ represents the power required for spectrum sensing operation, and τ is the transition time from sleep mode to active mode.

On the receiver side, the nodes detect the preamble transmission during spectrum sensing if the PU is not active. Hence, the expressions for energy drained in a single successful

and failed transmission on the j^{th} channel are given by

$$\mathcal{E}\_R^j_{succ} = \mathcal{E}^j_{ss} + P^j_{sw}\left\{(1-p)^m(\tau + T_s) + (1-p)^d(\tau + T_d\right\}\mathcal{P}_r, \tag{4.11}$$

$$\mathcal{E}\_R^j_{fail} = \mathcal{E}^j_{ss} + P^j_{sw}\left\{(\tau + T_s) + \left(1-(1-p)^d\right)(\tau + T_d)\right\}\mathcal{P}_r, \tag{4.12}$$

where $\mathcal{P}_r$ denotes the power drained in the receive mode. However, in CRB-MAC there are N eligible receivers that can forward the data packet. Therefore, the energy consumed in a single successful transmission in all possible cases where i nodes ($i \leq N$) successfully receive the packet without error is given by

$$\mathcal{E}\_R^j_{N\_succ} = \frac{\sum_{i=1}^{N}\binom{N}{i}\left[i\mathcal{E}\_R^j_{succ} + (N-i)\mathcal{E}\_R^j_{fail}\right]}{\sum_{i=1}^{N}\binom{N}{i}} \tag{4.13}$$

The energy consumed in a single transmission when all the receiver nodes fail to receive the packet without error is given by:

$$\mathcal{E}\_R^j_{N\_fail} = N \cdot \mathcal{E}\_R^j_{fail} \tag{4.14}$$

In case of a failed transmission, the sender node will retransmit the data. Hence, it is important to have a retransmission model for CRB-MAC. Let P_a denote the probability that a sender node will successfully transmit the packet after a failures. In CRB-MAC the sender node will stop retransmitting if atleast one of the receivers successfully receives the data frame. Thus, P_a is given by $P_a = (P^j_{fail})^{Na}(1-(P^j_{fail})^N)$. Let n_t be the random variable that represents the total number of transmissions until a success transmission. Since P_a is the Probability Mass Function (PMF) of n_t, the average number of retransmissions until success can be calculated as follows.

$$\begin{aligned}\chi = \mathbb{E}(n_t) &= \sum_{a=0}^{Z} a \cdot P_a = \sum_{a=0}^{Z} a \cdot (P^j_{fail})^{Na}(1-(P^j_{fail})^N) \\ &= \frac{(P^j_{fail})^N - (Z+1)(P^j_{fail})^{N(Z+1)} + Z(P^j_{fail})^{N(Z+2)}}{(1-(P^j_{fail})^N)},\end{aligned} \tag{4.15}$$

where Z represents the maximum number of retransmissions. In CR environments, the retransmissions also depend on channel availability. Hence, it is important to find the expected number of spectrum sensing events to find a vacant channel. Let χ^j_{ss} denote the

expected number of sensing events for transmitting over the j^{th} channel, which is given by

$$\chi_{ss}^{j} = \sum_{i=0}^{\infty} i \cdot (1 - P_{sw}^{j})^{i} P_{sw}^{j} = \frac{1 - P_{sw}^{j}}{P_{sw}^{j}} \tag{4.16}$$

Using the retransmission model, the total energy consumption for CRB-MAC (over the j^{th} channel) over a single hop is given by

$$\mathcal{E}_{CRB\_total}^{j} = \chi(\mathcal{E}\_T_{fail}^{j} + \mathcal{E}\_R_{N\_fail}^{j}) + \mathcal{E}\_T_{succ}^{j} + \mathcal{E}\_R_{N\_succ}^{j} + \chi_{ss}^{j}\mathcal{E}_{ss}^{j} \tag{4.17}$$

For a multi-hop scenario, the total energy consumption over H hops is given by $\sum_{h=1}^{H} \mathcal{E}_{CRB\_total}^{j}$.

4.3.3 Delay

Using the retransmission model described in the previous section, the single hop delay for CRB-MAC over the j^{th} channel is given by

$$\mathcal{D}_{CRB}^{j} = \chi \cdot (T_{pr} + T_{d} + T_{CW}) + \chi_{ss}^{j} \cdot T_{s} + T_{CI} \cdot (\chi_{ss}^{j} - 1) \tag{4.18}$$

For a multi-hop scenario, the end-to-end delay over H hops is given by $\sum_{h=1}^{H} \mathcal{D}_{CRB}^{j}$.

4.3.4 Reliability

In literature, *Packet Delivery Ratio* (PDR) is the most commonly used metric to quantify how reliably a protocol can deliver packets to the destination. The PDR is defined as the ratio of number of packets received to the total number of packets sent, and captures the fraction of packets actually delivered to the destination. However, the PDR is generally used in the context of routing protocols and hence implicitly evaluates the performance of underlying MAC protocol.

Analytically, the reliability for CRB-MAC (with N receivers) over the j^{th} channel is given by

$$\mathcal{R}_{CRB}^{j} = 100 \times \left(1 - (P_{fail}^{j})^{N(\chi+1)}\right) \tag{4.19}$$

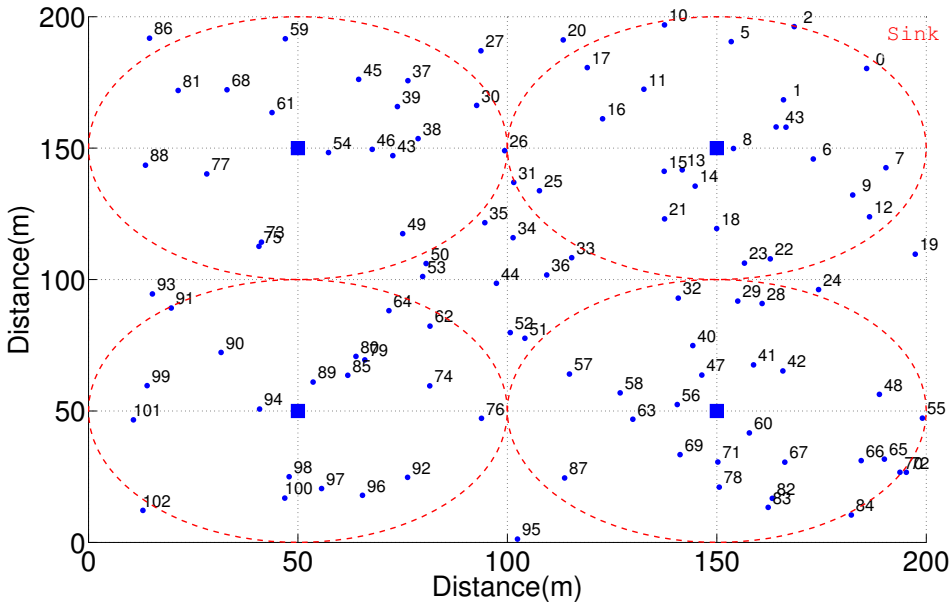


Figure 4.2: Sample simulated topology with Poisson distributed nodes (density = 0.4 nodes per unit area). Node ranks are also displayed. The filled squares and dotted circles represent the location and coverage area of PU transmitters respectively.

The multi-hop reliability for CRB-MAC over H hops is given by $\left[\mathcal{R}^{j}_{CRB}\right]^{H}$, which is essentially similar to PDR.

4.4 Performance Evaluation

In this section, we evaluate the single hop and multi-hop performance of CRB-MAC. We perform a MATLAB based simulation (with parameters given in Table 4.1) to validate the analytical models. A square region of side 200 meters is considered that is occupied by 4 PU transmitters. The secondary users are assumed to be Poisson distributed in the whole region with a mean density as shown in Fig. 4.2. Without loss of generality, we assume that RPL [10] is operating at the Network layer. Each node is assigned a rank (in RPL terminology, the rank represents a node's virtual position in the network with respect to the sink/gateway node (node 0 in our case)). For simplicity, we assume that the node's rank is dependent on its Euclidean distance from the sink node (i.e., the *objective function* is Euclidean distance in our case). The transmission radius of each node is set to 30 meters.

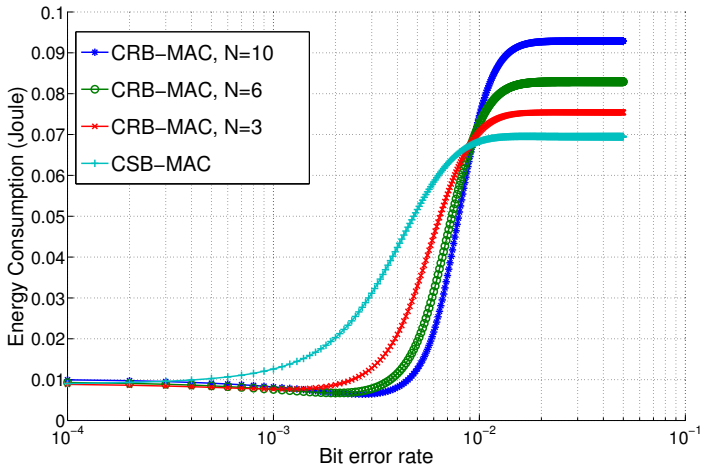
Table 4.1: Simulation Parameters for CRB-MAC

Parameter	**Value**
Detection probability threshold ($P_d^{'}$)	0.9
Probability of false alarm (P_f)	0.1
Channel bandwidth	200 KHz
PU received SNR (γ)	-15 dB
Busy state parameter of PU (μ_{ON})	2
Idle state parameter of PU (μ_{OFF})	3
Maximum Interference Ratio (IR_{max})	0.25
Spectrum sensing duration (T_s)	20 ms
CC2500 RF Transceiver Parameters	
Power drained in transmit mode ($\mathcal{P}_t$)	66.16 mW
Power drained in receive mode ($\mathcal{P}_r$)	70.69 mW
Power drained in spectrum sensing ($\mathcal{P}_s$)	65.83 mW
Checking interval (T_{CI})	144 ms
Preamble length (T_{pr})	144 ms
Transmission time of a data packet (T_d)	4 ms
Transmission time of one micro-frame (T_m)	40 μs
Transition time from sleep mode to active mode (τ)	88.4 μs

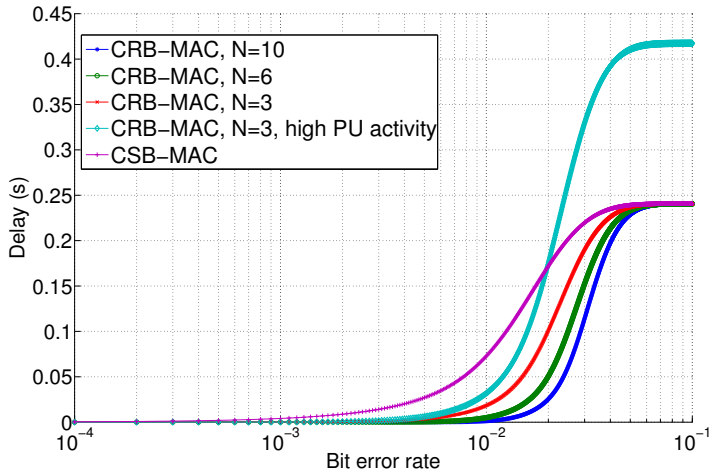
In addition, we assume that each node is equipped with *Texas Instruments* CC2500 Radio Transceiver whose parameters are also given in Table 4.1.

For comparison, we also implement a sender-based MAC protocol (1-hopMAC [117]) in CR environments (CSB-MAC). The analytical expressions for CSB-MAC are given in Appendix K.

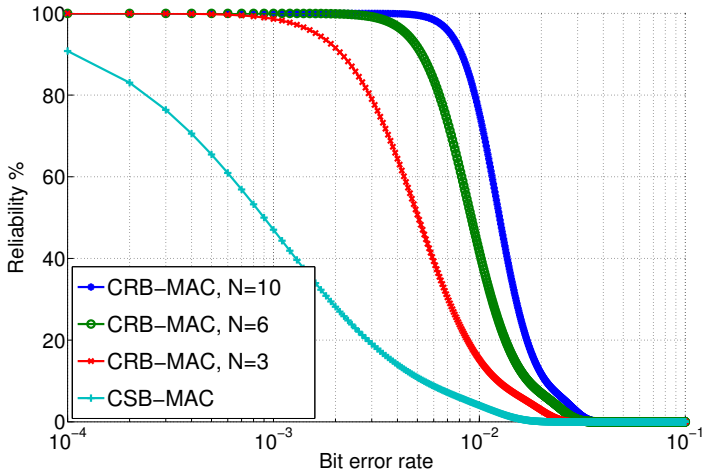
Fig. 4.3 evaluates the single hop performance of CRB-MAC (based on analytical models) for energy consumption (Fig. 4.3a), delay (Fig. 4.3b), and reliability (Fig. 4.3c) against the bit error rate (BER). First, we discuss the energy performance. In channels with rather low BER, CRB-MAC outperforms the CSB-MAC in terms of energy consumption. This is because, the number of retransmissions in CRB-MAC is less than CSB-MAC owing to multiple receivers involved in the forwarding process. That is why CRB-MAC is more resilient to variations in channel quality than CSB-MAC. In very poor channel conditions, CRB-MAC consumes more energy than CSB-MAC and the energy consumption increases with the number of receivers. The energy consumption reaches a saturation point when the maximum number of retransmissions (7 in our case) is reached. Therefore, the high energy



(a)



(b)

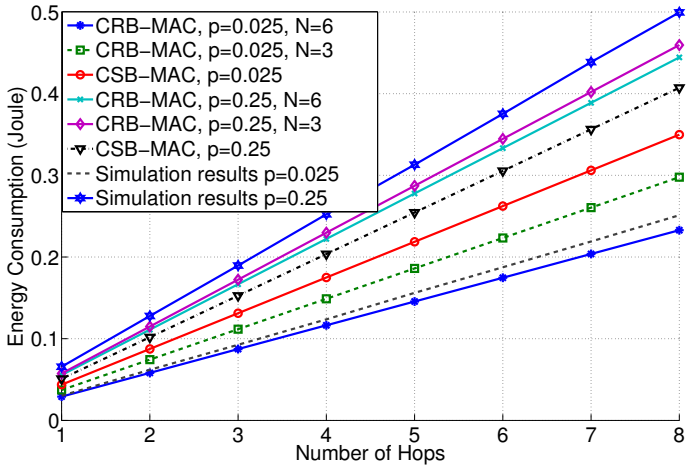


(c)

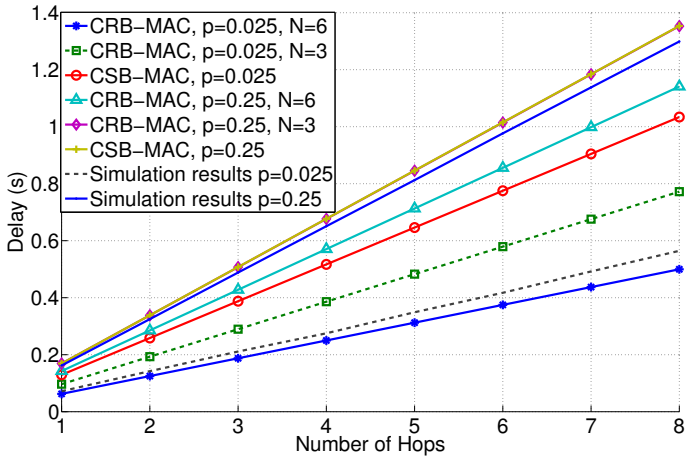
Figure 4.3: Single hop performance of CRB-MAC, (a) energy consumption, (b) delay, (c) reliability; N represents the number of receivers

consumption of CRB-MAC in poor channel conditions is primarily due to more receivers involved in the forwarding process. Hence, more energy is spent in the reception process that increases the overall energy consumption. Next, we discuss the delay performance. Since delay is dependent on the number of retransmissions, CRB-MAC outperforms CSB-MAC in terms of delay performance owing to fewer retransmissions. The delay reduces as the number of receivers increase because of higher probability of successful transmission. Moreover, the delay performance reaches a saturation point as the maximum number of retransmissions is reached. Note that a high PU activity ($P_b^j = 0.7$ in this case), with same number of receivers, further increases the delay due to more spectrum sensing events in order to find a vacant channel. Last, but not the least, we discuss the reliability performance. As expected, CSB-MAC (which relies on only one receiver) provides the lowest reliability. Whereas, CRB-MAC not only shows resiliency to channel quality variations but also provides much higher levels of reliability due to more receivers involved in the forwarding process.

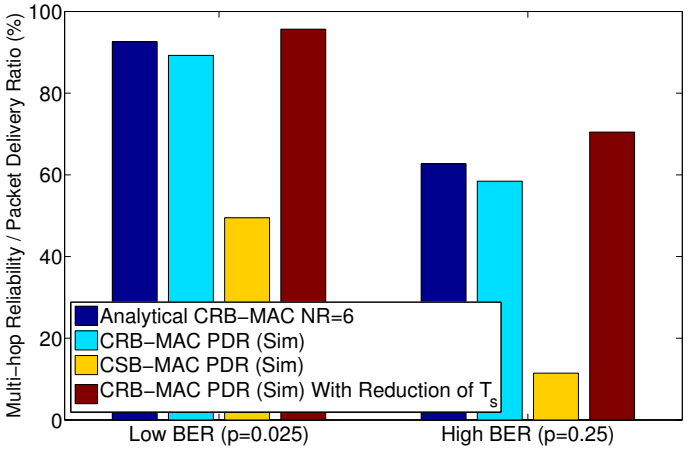
Next, we evaluate the multi-hop performance of CRB-MAC in both good (low BER with $p = 0.025$) and poor channel (high BER with $p = 0.25$) conditions. Figure 4.4a, 4.4b, and 4.4c respectively evaluate the multi-hop energy consumption, delay, and reliability of CRB-MAC. Simulation results for CRB-MAC are also presented. In simulations, we generate $10,000$ packets from different nodes and average the results for different performance metrics. Firstly, we discuss the energy performance. We note that the energy consumption increases with the number of hops, with CRB-MAC outperforming CSB-MAC in low BER conditions. In high BER conditions, the energy consumption of CRB-MAC increases due to higher energy consumption in the reception process as mentioned earlier. The simulation results follow the analytical results and hence validate the analytical modelling. Slight difference from analytical results is due to the fact that in simulations, nodes are randomly distributed and therefore, the number of receivers at each hop is not fixed (some nodes have fewer neighbors than others within the transmission range). Next, we see the delay performance. We note that CRB-MAC (with $N = 6$, and $N = 3$) outperforms CSB-MAC in terms of end-to-end delay (in both high and low BER scenarios) due to fewer retransmissions. Lastly, we discuss the reliability performance. We note that CRB-MAC provides better PDR (obtained through simulations) compared to CSB-MAC under both good and poor



(a)



(b)



(c)

Figure 4.4: Multi-hop performance of CRB-MAC, (a) energy consumption, (b) delay, (c) reliability; N = no. of receivers, p = bit error probability

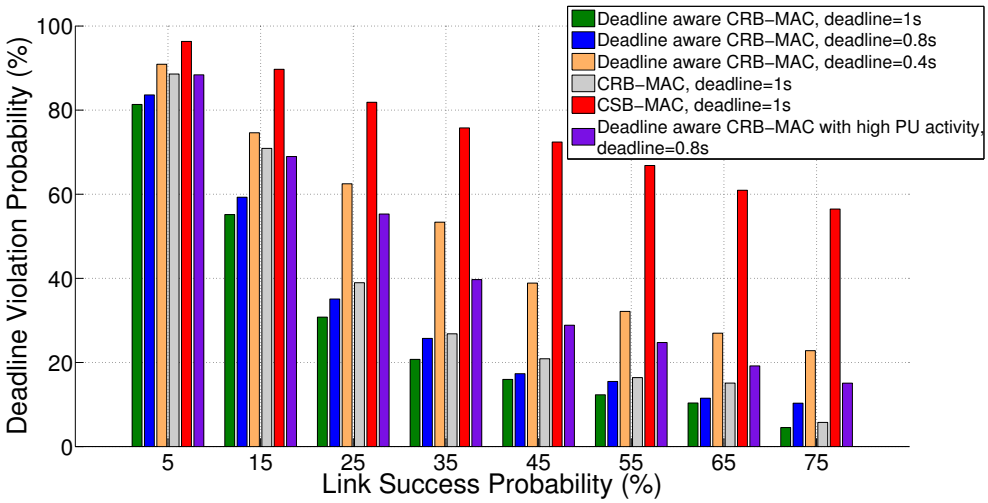


Figure 4.5: Simulation results for deadline violation probability against link success probability

channel conditions. Moreover, the multi-hop reliability obtained analytically (using (4.19)) is very close to the PDR which is obtained through simulations. We also note that a higher PDR is achieved by incorporating the performance enhancement technique for mitigating the degradation due to periodic spectrum sensing. The improvement in PDR results due to the reduction of sensing time by tracking the PU activity in the form of a moving window. Hence, fewer packets are dropped due to periodic spectrum sensing state of different nodes.

Lastly, we see the performance of deadline-aware CRB-MAC. The performance is compared in terms of *deadline violation probability* (DVP) which is calculated as the ratio of the number of packets dropped due to violation of deadline at intermediate hops to the total number of packets generated. The results in Fig. 4.5 show the simulation results for DVP against the link success probability. The DVP decreases as link success probability increases due to lesser retransmissions that increase the remaining lifetime of a packet at intermediate nodes. We note that the deadline-aware CRB-MAC achieves the best performance for any given deadline due to deadline-aware forwarding process. The CSB-MAC shows the worst performance as the forwarding process relies on a single receiver only which leads to higher number of retransmissions and consequently reducing the lifetime of a packet at intermediate nodes. We also note that a high PU activity ($P_b^j = 0.7$ in this case) increases the DVP as the delay in finding a vacant channel increases.

4.5 Summary and Concluding Remarks

M2M networks play an important role in the operation and management of smart grid. Due to challenging wireless environment, reliability and energy efficiency requirements for M2M devices become critically important. Considering the motivation of cognitive M2M and the requirement of M2M networks, we propose CRB-MAC which a receiver-based MAC protocol for cognitive M2M networks. CRB-MAC employs preamble sampling and opportunistic forwarding techniques to cater for high energy efficiency and reliability requirements of M2M networks. CRB-MAC is also enhanced for mitigating the performance degradation due to periodic spectrum sensing state of nodes in the network. Apart from this, a delay-budget based election process for packet forwarding is proposed in order to effectively handle delay sensitive traffic. Analytical and simulation results demonstrate that in lossy environments CRB-MAC generates less retransmissions and therefore, enhances the energy and delay performance. Moreover, high reliability can be provided by increasing the number of receivers (through dense deployments). Particularly, reliability improvement of up to 50% can be achieved compared to CSB-MAC protocol in lossy environments. Hence, CRB-MAC provides a viable solution for practical implementation in different areas of smart grid.

Chapter 5

Routing in Cognitive Radio Enabled Smart Grid AMI Networks

5.1 Introduction

One of the key elements of smart grid is the Advanced Metering Infrastructure (AMI) network, wherein multiple smart meters (located at customer premises) communicate with a local access point (meter concentrator) which is further connected to a Meter Data Management System (MDMS) that acts as a control center for storage, processing, and management of meter data in order to be used by different applications [33]. The AMI network can contribute in several ways to realize the vision of smart grid. For example, through the AMI network, utility providers can manage on demand power requirements, monitor power quality, identify anomalies, and regulate electricity usage (using dynamic pricing).

Depending upon the size of a utility, the number of smart meters in a network may vary from a few hundred to several thousand. Several communication technologies such as cellular, WiMAX, Power Line Communications (PLC), etc. are currently under consideration for AMI networks. However, there is no clear consensus by the community so far. Each of these technologies has its own pros and cons. For example, cellular networks are primarily optimized for conventional Human-to-Human (H2H) communication. Hence, radio resource management between H2H users and smart meters becomes challenging as both have different Quality-of-Service (QoS) requirements. Secondly, a large number of smart meters in a community can create traffic overload on the uplink random access channel. Thirdly, packet size for AMI type traffic can be much smaller than that of the signalling traffic, resulting in low efficiency [62]. Last, but not the least, cellular coverage penetration is an important issue that needs to be considered due to the variability in smart meter locations (e.g., some

meters may be installed at places such as garages, under the stairs, or may be present inside metal cages). Similar challenges exist for WiMAX based solutions. Apart from this, the security issues of WiMAX are still under investigation [120]. Moreover, utility providers are not comfortable with the fact that their data travels through a third party network; an issue which is common to both cellular and WiMAX based solutions. PLC appears to be an attractive solution due to the use of existing power grid infrastructure. However, the underlying communication medium will not be available in case of power outage which is a serious issue. Moreover, in some parts of the world (e.g., Norway) regulatory authorities have banned the use of PLC due to possible detrimental effect on military HF radio communications [121].

A practical solution is to deploy a static multi-hop wireless mesh network connecting a large number of smart meters which in turn is connected to a gateway (concentrator). This solution is particularly attractive as it scales well with the size of the AMI network. In addition, the utility provider has complete control over the infrastructure. It should be noted that although smart meters are static, the wireless link between any arbitrary pair of smart meters is generally unstable due to fading and interference effects. Therefore, the AMI network requires proper routing functionalities for reliable and low latency delivery of data for different applications.

RPL (Routing Protocol for Low Power and Lossy Networks) [10] is a routing protocol that has been recently standardized by IETF and intends to support a variety of applications including building automation, healthcare, urban sensor networks, industrial monitoring, etc. RPL is currently under active investigation in the research community. Moreover, it is attracting a lot of attention for AMI mesh networks in smart grids (e.g., see [122–126]) and is expected to be the standard routing protocol for AMI applications.

As discussed in Section 2.2 and 3.1.1, the use of cognitive radio (CR) technology is extremely promising for M2M communications in general and smart grid communications in particular. Recently, a number of studies (e.g., see [78–83]) have been presented on different smart grid related platforms regarding the application of cognitive radio for smart grid communication.

Against this background, our objective in this chapter is to enhance RPL for CR

enabled AMI networks. To the best of our knowledge, the adaptability and application of RPL in CR enabled AMI networks has not been studied before. We enhance basic RPL with novel modifications especially tailored for CR environments. We develop an opportunistic forwarding approach to meet the utility requirements of secondary network (cognitive AMI network) along with protecting the primary users (PUs). The enhanced protocol is termed as CORPL (Cognitive and Opportunistic RPL). The rest of the chapter is organized as follows. Section 5.2 presents an overview of RPL. In Section 5.3, we discuss the challenges for any routing protocol in CR environments. CORPL considers these challenges as design objectives. Section 5.4 presents the CORPL framework followed by the performance evaluation in Section 5.5. Finally the chapter is summarized and concluded in Section 5.6.

5.2 Overview of RPL

RPL is a distance-vector and a source routing protocol. The key aspect of RPL is to maintain network state information using one or more *Directed Acyclic Graphs* (DAGs). A DAG is a directed graph wherein all edges are oriented in such a way that no cycles exist. Each DAG created in RPL has a root node which acts as a gateway. Each node (client node) in the DAG is assigned a rank that is computed on the basis of an *objective function*. The rank monotonically increases in the downward direction (DAG root has the lowest rank) and represents a node's virtual position to other nodes with respect to the DAG root. A node in DAG can only be associated with other nodes having same or smaller rank compared to its own rank in order to avoid cycles. RPL does not specify any particular objective function for DAG rank computation.

In order to construct a DAG, the gateway broadcasts a control message called *DAG Information Object* (DIO) containing relevant network information including the DAGID to identify the DAG and the rank information along with the objective function for rank computation. Any node that receives the DIO message and wants to join the DAG should add the DIO sender to its parent list, compute its own rank according to the objective function, and forward the DIO message with the updated rank information. When a node already

associated with the DAG receives another DIO message, it can discard the DIO message (according to some criteria specified by RPL), process the DIO message to maintain its position in existing DAG, or improve its position by obtaining a lower rank according to the objective function. Once the DAG is constructed, each node will be able to forward any inward traffic (destined to the gateway) by choosing its most preferred parent as the next hop node.

RPL also specifies a methodology for outward traffic (gateway to client node) through *Destination Advertisement Object* (DAO) control message which is unicast in the upward direction. The intermediate nodes record the reverse path information and thus a complete outward path is established from the gateway to the client node.

To maintain a DAG, each node periodically generates DIO messages triggered by the *trickle timer* [127] which optimizes the message transmission frequency based on network conditions. The frequency is increased in case of inconsistent network information and decreased in case of stable network conditions. For more information on RPL, the interested reader is referred to comprehensive surveys in [48] and [128].

5.3 Routing Challenges in CR Environment

Spectrum sensing is a key aspect of any CR environment. Nodes periodically monitor the current channel for PU activity before using it for transmission. During this interval (sensing time), nodes are not involved in forwarding data packets and therefore, the multi-hop network is virtually disconnected at the node that is engaged in spectrum sensing. Hence, the routing algorithm should explicitly account for the spectrum sensing state of different nodes.

The secondary network operation must ensure protection for both PU transmitters and PU receivers i.e., temporal and spatial protection respectively. The latter is particularly important for those PU applications where the transmission is uni-directional (e.g., TV broadcast). The protection to the PU transmitter is subject to accurate detection of the PU activity. On the other hand, PU receivers are difficult to detect and can be easily affected by the transmission from neighboring CR users. Therefore, the network layer should provide explicit protection to PU receivers by avoiding regions where such users might be

present [129].

The protection provided to PUs results in a performance trade-off for the secondary network. Hence the routing protocol must optimize the operation for both primary and secondary networks depending upon the level of protection for the former and the Quality-of-Service (QoS) requirements of the latter.

5.4 CORPL Framework

In this section, we describe the framework of our enhanced RPL protocol for CR environments i.e., CORPL. The objective of CORPL is to retain the DAG based approach of RPL and at the same time introduce novel modifications to allow its application in CR environments.

5.4.1 CORPL Overview

To address the afore mentioned challenges, we develop an opportunistic forwarding approach [130] that consists of two key steps: selection of a forwarder set i.e., each node in the network selects multiple next hop neighbors, and a coordination scheme to ensure that only the best receiver of each packet forwards it (unique forwarder selection). It has been shown that the opportunistic forwarding approach improves the end-to-end throughput and reliability (by exploiting the inherent characteristics of wireless channel) of the network; the latter being an important requirement for lossy networks.

A key challenge in opportunistic forwarding is the selection of forwarder set. CORPL takes advantage of the existing parent structure of RPL that requires at least one backup parent besides the default parent[1]. In CORPL, each node maintains a forwarder set such that the forwarding node (next hop) is opportunistically selected. The creation of forwarder set is elaborated upon later. CORPL uses a cost function approach to dynamically prioritize the nodes in the forwarder set. Moreover, CORPL uses a simple overhearing based coordination scheme to ensure a unique forwarder selection.

[1]In RPL the backup parent ignores the transmission and the packet is forwarded through the default parent only.

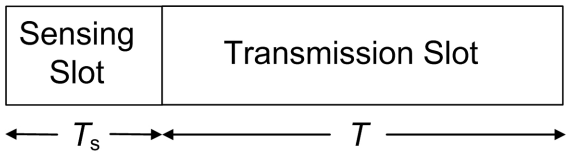


Figure 5.1: MAC frame structure in a CR network

CORPL takes advantage of the opportunistic forwarding approach to support high-priority delay sensitive alarms that need to arrive at the gateway before a given deadline as well as to select paths with minimum interference to PU receivers. The PU transmitter protection is ensured through optimal transmission time for the secondary network subject to an interference constraint. This will be discussed in detail later.

Since nodes engaged in spectrum sensing cannot receive/forward packets, the network performance is degraded in terms of end-to-end throughput, latency, and packet loss ratio. CORPL utilizes two different techniques to improve overall network performance under spectrum sensing state of different nodes.

5.4.2 Network Model

We consider a static multi-hop wireless AMI network consisting of different smart meters and a gateway node (meter concentrator). We assume that the smart meters are CR enabled[2]. Each smart meter (node) is equipped with a single radio transceiver that can be tuned to any channel in the licensed spectrum. We assume N stationary PU transmitters (and hence N available channels) with known locations and maximum coverage ranges. As in Section 4.2.2, the PU (transmitter) activity model for the j^{th} channel is given by a two state independent and identically distributed (i.i.d.) random process such that the duration of busy and idle periods is exponentially distributed with a mean of $\frac{1}{\mu^j_{ON}}$ and $\frac{1}{\mu^j_{OFF}}$, respectively. Let S^j_b denote the state that the j^{th} channel is busy (PU is active) with probability $P^j_b = \frac{\mu^j_{OFF}}{\mu^j_{ON}+\mu^j_{OFF}}$, and S^j_i denote the state that the j^{th} channel is idle with probability P^j_i, such that $P^j_i + P^j_b = 1$. We assume that a node employs energy detection

[2]It should be noted that due to resource constrained nature of smart meters, there is a need of developing low cost dynamic spectrum access solution for (ad-hoc type) cognitive AMI networks. However, this is beyond the scope of this chapter.

Figure 5.2: Rank computation based on ETX. The default parent for node 5 is node 3 owing to a smaller rank compared to nodes 2 and 4. An example forwarder list for *Class A* routes (obtained using (5.8)) is also shown along with the timeline of coordination scheme. Note that a node who fails to receive the ACK will forward the frame as well.

technique [99] (during spectrum sensing period) for primary signal detection wherein it compares the received energy (E) with a predefined threshold (σ) to decide whether the j^{th} channel is occupied or not i.e.,

$$Sensing\ Decision = \begin{cases} S_b^j & \text{if} \quad E \geq \sigma \\ S_i^j & \text{if} \quad E < \sigma \end{cases} \tag{5.1}$$

The two principle metrics in spectrum sensing are the detection probability (P_d), and the false alarm probability (P_f). A higher detection probability ensures better protection to incumbents, whereas a lower false alarm probability ensures efficient utilization of the channel. As per [98], false alarm and detection probabilities for the j^{th} channel can be expressed as follows.

$$P_f^j = Pr\left\{E \geq \sigma \mid S_i^j\right\} = \frac{1}{2} erfc\left(\frac{1}{\sqrt{2}} \frac{\sigma - 2n_j}{\sqrt{4n_j}}\right), \tag{5.2}$$

$$P_d^j = Pr\left\{E \geq \sigma \mid S_b^j\right\} = \frac{1}{2} erfc\left(\frac{1}{\sqrt{2}} \frac{\sigma - 2n_j\left(\gamma_j + 1\right)}{\sqrt{4n_j\left(2\gamma_j + 1\right)}}\right), \tag{5.3}$$

where $erfc(\cdot)$ is the *complementary error function*, and γ_j and n_j denote the signal-to-noise ratio (SNR) of the primary signal and the bandwidth-time product for the j^{th} channel respectively.

The MAC frame structure in a CR network consists of a sensing slot (T_s) and a transmission slot (T) as shown in Fig. 5.1. In periodic spectrum sensing scenarios, there is a possibility of causing harmful interference to PUs due to imperfect spectrum sensing in realistic conditions. This interference is quantified in terms of *Interference Ratio* (IR), defined as the expected fraction of ON duration of PU transmission interrupted by the transmission of secondary users and is given for the j^{th} channel as follows [100].

$$IR_j = \left(1 - P_d^j\right) P_b^j + P_i^j \left(1 - P_f^j\right) + e^{-\mu T}\left(P_f^j - P_d^j\right), \tag{5.4}$$

where $\mu = \max(\mu_{ON}^j, \mu_{OFF}^j)$. We assume that the nodes in our network employ optimal transmission time that maximizes the throughput of the secondary network subject to an

interference constraint i.e., $IR_j \leq IR^j_{max}$, where IR^j_{max} denotes the maximum tolerable interference ratio on the j^{th} channel. This transmission time for the j^{th} channel is given as follows.

$$T_j = \mu^{-1}\left[\ln P^j_i - \ln\left(P^j_i P^{'}_d + P^j_b(1-P^{'}_d) - IR^j_{max}\right) + \ln(2P^{'}_d - 1)\right], \tag{5.5}$$

where $P^{'}_d$ is the detection probability threshold, defined as the detection probability at SNR level as low as γ_{min}, where γ_{min} is specified by the regulator.

We assume that the AMI network comprises of two types of traffic: low priority monitoring data (that can be considered as best-effort) and high priority delay sensitive[3] information (that has an associated deadline).

5.4.3 Protocol Description

Since we want to retain the DAG structure of RPL, the DAG construction process in CORPL follows a similar procedure as explained earlier. After detecting a vacant channel, the gateway node transmits a DIO message. We use *Expected Transmission Count* (ETX) [119] as the default metric for rank computation, which is frequently used in lossy networks. The ETX of a link from node a to node b is given by $E_{ab} = 1/p_{ab}$, where p_{ab} is the probability of node b receiving a transmission from node a. The ETX of a link will be measured and updated continuously once the link starts to carry data traffic. The rank computation method for a node joining the DAG is illustrated in Fig. 5.2. Due to periodic spectrum sensing by each node, the DAG convergence time (defined as the time taken by the set of nodes to obtain topological information and become part of the DAG) will increase due to higher packet loss ratio as explained earlier.

Next we describe the procedure of constructing the forwarder set for opportunistic forwarding. It should be noted that each node in CORPL has a default parent (like RPL) which has been selected based on ETX. The forwarder set is constructed in such a way that the forwarding nodes are within the transmission range of each other. During the DIO transmission, each node also reports some additional information using the Option field of the DIO message[4]. Each node updates the neighborhood information through the DIO

[3]A fundamental objective in smart grid is to balance the supply and demand through precise information of power load obtained via smart meters. Hence, delay sensitive traffic is an integral part of AMI networks.

[4]One option field is limited to 7 bytes, such that 1 byte is allocated to 'Option Type', 1 byte is allocated

message transmission. Based upon the neighborhood information, each node dynamically prioritizes its neighbors in order to construct the forwarder list[5] . The priorities are assigned according to a cost function. Since the construction of forwarder list incurs overhead, the size of forwarder set is limited to a maximum of M neighbors. When a node does not hear from its neighbor for a predefined time interval, its corresponding entry in the forwarder list is deleted. Similarly, the forwarder list is updated if a node having a better cost appears.

The cost function to prioritize the nodes in the forwarder set depends on the routing class. CORPL considers two different routing classes. The first class (*class A*) assigns a greater importance to PU receiver protection, whereas in the second class (*class B*), end-to-end latency is the key consideration for supporting high priority delay sensitive alarms. These two classes of protocols are explained as follows.

In order to reduce interference to PU receivers (which can be present anywhere in the coverage area of PU transmitters), the routes for the secondary network should be selected such that they pass through regions of minimum coverage overlap with the PU transmission coverage. A node k calculates the fractional area of its transmission coverage under the coverage of j^{th} PU transmitter as $C_{kj} = C'_{kj}/\pi r_k^2$, such that C'_{kj} is given by

$$\begin{aligned} C'_{kj} = r_k^2 \cos^{-1}\left(\frac{d_{kj}^2 + r_k^2 - R_j^2}{2d_{kj}r_k}\right) + R_j^2 \cos^{-1}\left(\frac{d_{kj}^2 + R_j^2 - r_k^2}{2d_{kj}R_j}\right) \\ - 0.5 \times \sqrt{\left\{(R_j + r_k)^2 - d_{kj}^2\right\}(d_{kj} + r_k - R_j)(d_{kj} - r_k + R_j)}, \end{aligned} \tag{5.6}$$

where R_j and r_k denote the coverage radii of the j^{th} PU transmitter and the k^{th} node respectively, and d_{kj} is the distance between the two.

For delay sensitive alarms, the node must find the next-hop that guarantees the deadline. If the deadline has elapsed, the packet will be dropped. The node assumes that the time before the deadline can be uniformly shared among the nodes in the route. The *delay budget* (DB) for the transmission is given by

$$DB = \frac{deadline(P) - t}{d(k)}, \tag{5.7}$$

to 'Option Length', and 5 bytes are allocated for 'Option Data'. We assume that 5 bytes are sufficient for including neighbor address along with necessary neighborhood information.

[5]The forwarder list refers to the arrangement of nodes in the forwarder set according to their respective priorities.

where $deadline(P)$ is the deadline associated with the packet P, t is the current time, and $d(k)$ is the hop distance between the k^{th} node and the DAG root. When a packet is at node k, the delay before the packet is correctly transmitted to the next hop depends on: (a) delay until a vacant channel is found (t_1) and (b) the average delay until the next hop correctly receives the packet (t_2). While t_1 depends on PU activity and spectrum sensing outcome, t_2 is characterized by the MAC layer and can be estimated through the packet delivery ratio. The node that provides the highest margin for delay budget i.e., ($DB \geq t_1 + t_2$) will be given the highest priority in the forwarder list.

The nodes in the forwarder list are prioritized according to a cost function based on the received neighborhood information. A node k calculates the cost for a node i in its forwarder set as follows.

$$\delta_i = \omega_1 \cdot C_i + \omega_2 \cdot E_{ki} + \omega_3 \cdot DBM_i, \tag{5.8}$$

where $C_i = \sum_{j=1}^{N} C_{ij}$ is the net overlapping area of i^{th} node with all PU transmitters, E_{ki} is the ETX of the link between nodes k and i, DBM_i accounts for the delay budget margin provided by the i^{th} node, and ω_1, ω_2, and ω_3 are design parameters such that $\omega_1 + \omega_2 + \omega_3 = 1$. For *class A* routes, $\omega_1 \gg \omega_2, \omega_3$ and the node with minimum cost has the highest priority. For *class B* routes, $\omega_3 \gg \omega_1, \omega_2$ and the node with the highest cost has the highest priority. The cost function also includes a weightage for ETX which is a link quality indicator. It has been shown that the cooperative gain of opportunistic forwarding becomes less significant when the inter-forwarder link success probabilities are low [119]. Thus, it is important to consider the effect of ETX in selecting the forwarding nodes.

CORPL requires some modifications at the MAC layer as well. In CORPL setup, the MAC layer adds the addresses of the nodes in forwarder list to the MAC header of the frame. The receiving nodes (nodes in the forwarder set) extract the address information (added on top of the standard header) by decoding MAC header. A node obtains the priority information by checking the location of its address in the MAC header (e.g., if its address is in the first address location of the header, it has the highest priority in the forwarder list).

In CORPL, the default parent has the highest priority for best-effort traffic. However, for *class A* and *class B* routes, the default parent is also considered in the forwarding set.

If a better node (with lower or higher cost for *class A* or *class B* respectively) is available, a special flag is set in the header (6LoWPAN packet header) of forwarding packet which indicates that the packet is not intended for the default parent. In this case the default parent follows a similar procedure as described earlier for any other receiving node. A pseudocode for packet forwarding in CORPL is given as Algorithm 1.

In order to ensure unique forwarder selection, CORPL employs a simple overhearing-based coordination scheme based on the acknowledgement (ACK) frames. This is illustrated in Fig. 5.2. If the special flag is not set, the default parent forwards the data to the next hop and generates an ACK. This ACK is captured by the nodes in the forwarder set (recall that the nodes forming the forwarder set are within transmission range of each other). If the default parent fails to forward the frame within a timeout period (no ACK is received), the node with the next highest priority forwards it. In case of *class A* or *class B* routes, the highest priority node forwards the data by default and in case it fails to forward, the second highest priority node forwards it with the same technique. It should be noted that this approach has an associated probability of erroneous forwarding of the same frame by multiple forwarding nodes. Thus, we define *coordination overhead* as the probability of a node in the forwarder set retransmitting a frame when any other node has already forwarded it to next hop. The coordination overhead (O_c) for a node a whose parent set is indicated by $\mathcal{P}_s^a$ can be calculated as follows.

$$O_c^a = \sum_{b=1}^{|\mathcal{P}_s^a|} p_{ab} \cdot EC_{bg} \cdot \prod_{r=1}^{b-1}(1 - p_{ar}), \tag{5.9}$$

where EC_{bg} is the path cost from node b to the gateway node. Since the rank computation is based on ETX, we assume the path cost in terms of ETX for calculating O_c.

The total path cost to reach the gateway node from a node b with a parent set $\mathcal{P}_s^b$ depends on the cost of opportunistic forwarding to its parent set and the remaining path cost of node b's parent set [131], which is given by

$$EC_{bg} = \frac{1 + \left\{ X_1 p_{b1} + \sum_{j=2}^{|\mathcal{P}_s^b|} X_{bj} p_{bj} \cdot \prod_{n=1}^{j-1}(1 - p_{bn}) \right\}}{1 - \prod_{j \in \mathcal{P}_s^b}(1 - p_{bj})}, \tag{5.10}$$

where it is assumed that the nodes in $\mathcal{P}_s^b$ are sorted by their cost (in terms of ETX) to the

Algorithm 1: Opportunistic Forwarding in CORPL

$\mathcal{K}^n = \{k_1, k_2, \cdots, k_M\} \longrightarrow$ forwarder set of node n
$k^{'} \in \mathcal{K}^n \longrightarrow$ default parent of node n
$\mathcal{Q}^n_{class\ A} = \{\cdot\}_{1\times M}$, $\mathcal{Q}^n_{class\ B} = \{\cdot\}_{1\times M}$
for $i = 1 : \mid \mathcal{K}^n \mid$ **do**
  if *class A* **then**
    calculate $\delta_i \mid \omega_1 \gg \omega_2, \omega_3$ as per (5.8)
    $\mathcal{Q}^n_{class\ A}(i) \leftarrow \delta_i$;
    else if *class B* **then**
      calculate $\delta_i \mid \omega_3 \gg \omega_1, \omega_2$ as per (5.8)
      $\mathcal{Q}^n_{class\ B}(i) \leftarrow \delta_i$;
    end
  end
end
sort $\mathcal{Q}^n_{class\ A}$ and $\mathcal{Q}^n_{class\ B}$ in ascending order
(Forwarding Rules)
if *incoming packet belongs to class A* **then**
  highest priority next hop = $k^* \in \mathcal{K}^n$ with cost = $\mathcal{Q}^n_{class\ A}(1)$
  else if *incoming packet belongs to class B* **then**
    highest priority next hop = $k^* \in \mathcal{K}^n$ with cost = $\mathcal{Q}^n_{class\ \boldsymbol{A}}(M)$
  end
  else
    highest priority next hop = $k^{'} \in \mathcal{K}^n$
  end
end

gateway node i.e., $X_1 < X_2 < \cdots < X_{|\mathcal{P}_s^b|}$. Note that the second term in the numerator accounts for the probability of a data packet being received by a particular node in $\mathcal{P}_s^b$ and not being received by any node with a lower cost to reach the gateway node, whereas the denominator accounts for the probability that at least one node in $\mathcal{P}_s^b$ has received the packet.

CORPL employs two different techniques for mitigating the performance degradation due to spectrum sensing. The first technique improves the performance through gathering sensing schedule information of the neighboring nodes. During DIO message transmission, each node also appends the following information: (a) time left before the node starts the next round of spectrum sensing, (b) interval between two successive spectrum sensing events, and (c) timestamp. A receiving node maintains this information along with the forwarder list. Therefore, a node knows when its neighboring nodes will undertake spectrum sensing and for how long. This is particularly important for delay sensitive traffic. Nodes which are unable to forward packets due to spectrum sensing, and hence provide a lower delay budget margin can be avoided by assigning a lower priority in the forwarder list.

The second technique improves performance by decreasing the spectrum sensing time. Reduction of sensing time is possible when a node is situated in region of low PU activity, and hence the number of channel changes that occur over time is small [118]. Initially the sensing time is set to maximum value i.e., $T_s = T_s^{max}$ for a fixed missed detection probability ($P_m = 1 - P_d$). The sensing time is decreased over time (by tracking the PU activity and establishing the fact that the node is located in region of low PU activity) according to the following relation: $T_s^{new} = T_s - \varphi \cdot \Delta_s$, where Δ_s is the step size, given by $\Delta_s = 0.5 \times T_s$ and φ is a constant which is obtained from the gradient of sensing time versus the missed detection probability curve (see [118] for more details). When successive missed detection events occur, the node increases the sensing time with similar step size.

It should be noted that traffic in AMI networks is mostly inward (from nodes to gateway), therefore CORPL primarily focuses on inward traffic. The outward traffic, which is rare, follows the standard reverse path recording methodology [48] using DAO messages as described in the RPL standard.

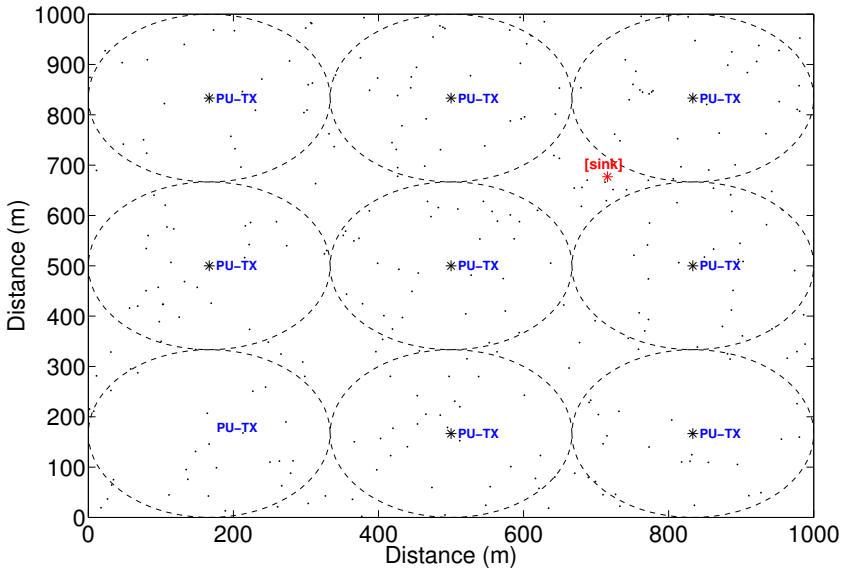
Table 5.1: Simulation Parameters for CORPL

Parameter	**Value**
Path loss model	$128.1 + 37.6\log_{10}(r)$, r in km, carrier freq = 2 GHz
Standard deviation of shadowing	8 dB
Detection probability threshold ($P_d^{'}$)	0.9
Probability of false alarm (P_f)	0.1
Channel bandwidth	200 KHz
PU received SNR (γ)	-15 dB
Busy state parameter of PU (μ_{ON})	2
Idle state parameter of PU (μ_{OFF})	3
Maximum Interference Ratio (IR_{max})	0.25
Size of forwarder set (M)	5
Size of DIO message including options	28 bytes

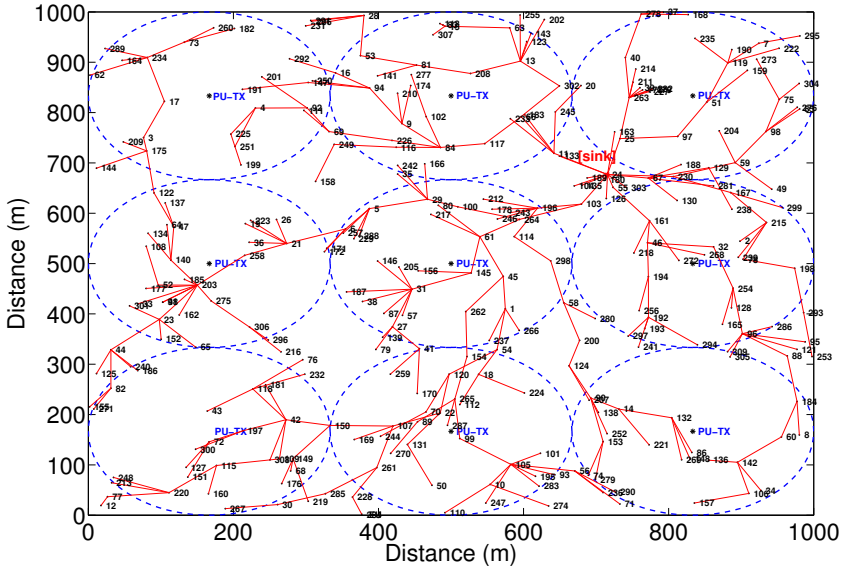
5.5 Performance Evaluation

In this section, we evaluate the performance of CORPL in different scenarios. We implement CORPL in MATLAB with the topology as shown in Fig. 5.3. Other simulation parameters are given in Table 5.1. We consider a square region of side 1000 meters that is occupied by 9 PU transmitters. The secondary users are assumed to be Poisson distributed in the whole region with a mean density as shown. We consider a frequency selective Rayleigh fading channel between any two nodes, where the channel gain accounts for small scale Rayleigh fading, large scale path loss, and shadowing. For performance comparison, we also implement RPL in CR environments.

First we investigate the impact of spectrum sensing on the overall performance. Fig. 5.4 shows the average DAG convergence time against the spectrum sensing time. The results are averaged over 100 iterations and represented in the form of a box plot. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentile, and the whiskers extend to the most extreme data points not considered as outliers. The link outage probability (LOP) is set to 20%. We note that the DAG convergence time increases as sensing time increases due to the fact that DIO messages are dropped with a higher probability (as nodes spend more time in spectrum sensing state). Hence, a large number of DIO message retransmissions contribute to a higher DAG convergence time.



(a) PU transmitters and Poisson distributed secondary nodes



(b) DAG construction

Figure 5.3: Simulated network topology. The circles represent the coverage area of PU transmitters. In order to have realistic number of secondary nodes, the density is kept low (node density $= 3 \times 10^{-4}$ nodes per unit area in this case). Nodes are connected in the form of a DAG where numbers represent node IDs

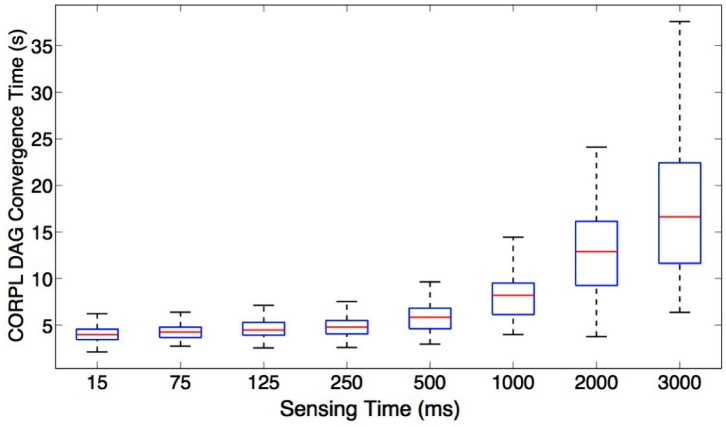


Figure 5.4: Spectrum sensing time against the average DAG convergence time over 100 iterations (node density = 3×10^{-4} nodes per unit area)

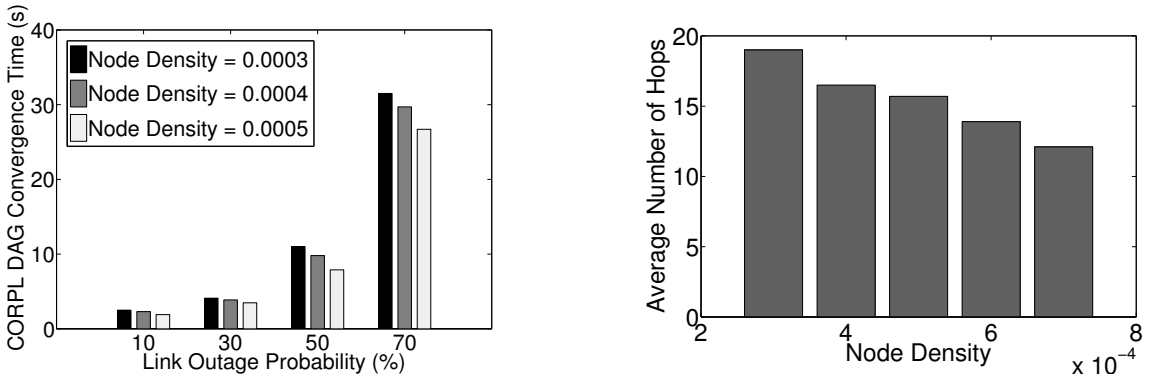


Figure 5.5: (a) DAG convergence time against LOP, (b) Average no. of hops towards gateway for different node densities

Similarly, the DAG convergence time increases as the LOP increases due to higher link layer retransmissions as shown in Fig. 5.5a. Note that the DAG convergence time reduces as the node density increases. This is because a higher density results in faster dissemination of network information owing to more nodes in the coverage range. Moreover, the probability of a node associating with a lower ranked parent increases which ultimately improves the DAG convergence time by reducing the number of hops towards the gateway as shown in Fig. 5.5b.

Next, we evaluate the performance in terms of *Packet Delivery Ratio* (PDR), which is defined as the ratio of the number of packets received to the total number of packets generated. PDR captures the fraction of packets sent by different nodes that are actually

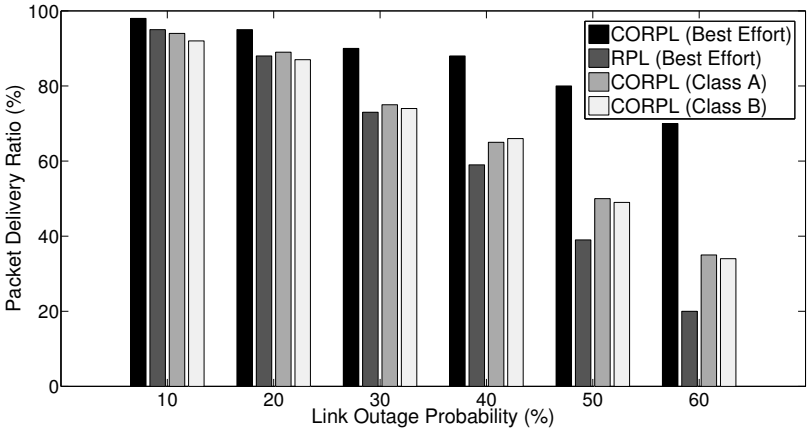


Figure 5.6: PDR performance comparison for different protocols (node density = 3×10^{-4} nodes per unit area)

delivered to the gateway. We generate 10,000 packets (packet size = 100 bytes) from different nodes and calculate the average PDR for different scenarios as shown in Fig. 5.6. It is evident from the results that CORPL outperforms RPL, where traffic is forwarded through the default parent only. The performance gain is significant under poor channel conditions (high LOP). CORPL utilizes the diversity of routes and hence improves the PDR by reducing retransmissions. For best-effort traffic in CORPL, ETX is the only factor in ranking the nodes in the forwarder set. Hence, the PDR for best-effort traffic is higher than *class A* and *class B* routes which assign a relatively less weightage to ETX.

We also evaluate the class specific performance of CORPL. The results in Fig. 5.7 evaluate the *Deadline Violation Probability* (DVP) for delay sensitive alarms in different scenarios. The DVP increases as the LOP increases due to higher link layer retransmissions that decrease the remaining lifetime of a packet at the intermediate nodes and therefore, the packet is dropped before reaching the gateway. CORPL (*class B*) provides enhanced performance compared to RPL as the next hop is opportunistically selected in the former by assigning higher priority to nodes providing higher delay budget margin. This is unlike RPL where the default parent may not always provide enough delay budget margin. Moreover, a higher node density reduces the DVP by reducing the number of hops towards the gateway as shown in Fig. 5.5b.

We evaluate the level of protection for PU receivers in terms of *Collision Risk Factor*

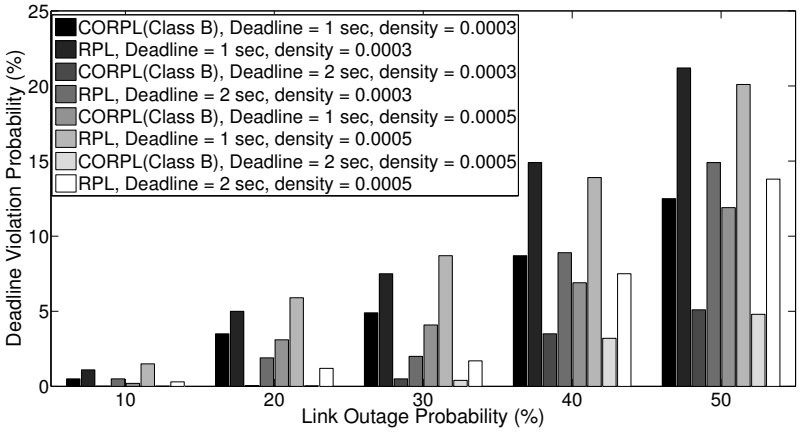


Figure 5.7: Deadline Violation Probability for different scenarios (averaged over $10,000$ packets from different nodes, the order of legend applies left to right)

(CRF), which is defined as the ratio of colliding transmissions to the total number of secondary node transmissions at the PU receivers. Hence CRF depends on the PU transmitter activity and the coverage overlap between secondary nodes and PU transmitters. As seen by the results in Fig. 5.8, CORPL (*class A*) reduces the chances of collision to PU receivers by up to 50% under both low and high PU transmitter activity. Note that the CRF increases with increased PU activity as well as with larger secondary node transmission range due to higher collision probability with PU receivers.

CORPL employs two different techniques for mitigating the performance degradation due to spectrum sensing, which have been evaluated in Fig. 5.9 by calculating the average DAG convergence time. *Techniques A* and *B* respectively refer to gathering sensing schedule information and reducing the sensing time under low PU activity, whereas the standard method employs no enhancement technique. Both techniques improve the DAG convergence time. However, the highest improvement is achieved through *technique B* where the sensing time is reduced over time by tracking the PU activity in the form of a moving window. The abscissa in Fig. 5.9 refers to T_s^{max} for *technique B*, using which the step size is calculated as described earlier. Both techniques will also enhance the DAG maintenance phase by reducing the number of packets dropped due to periodic spectrum sensing state.

Lastly, we evaluate the coordination overhead (O_c) of CORPL. In simulations, it is estimated as the ratio of the number of duplicate packets to the total number of packets

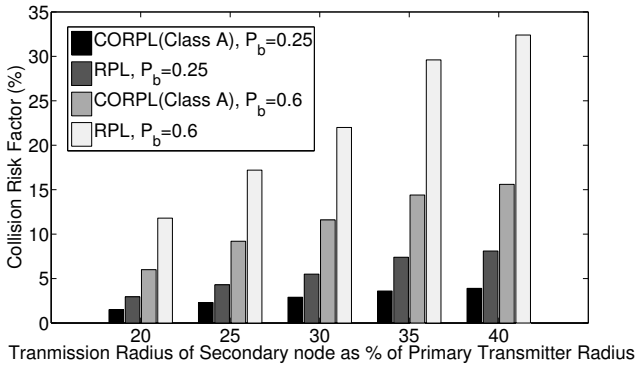


Figure 5.8: Collision Risk Factor against secondary nodes transmission radii (results are averaged over $10,000$ packets from different nodes, node density $= 3 \times 10^{-4}$ nodes per unit area)

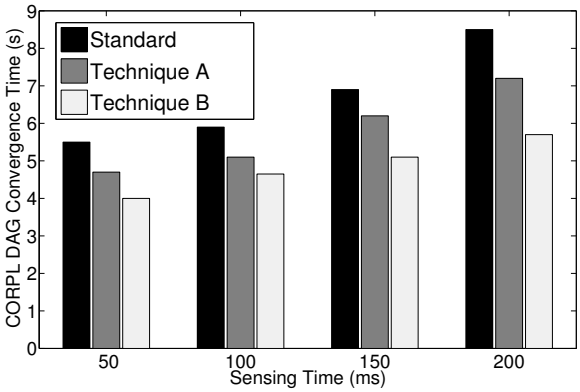


Figure 5.9: DAG convergence time for different performance enhancement techniques (LOP = 20%, results are averaged over 100 iterations, node density $= 3 \times 10^{-4}$ nodes per unit)

received at the gateway node. The results in Fig. 5.10 show the trend of O_c against the size of forwarder set (M) for different values of LOP. We note that O_c increases as LOP increases due to the fact that the probability of a node (in the forwarding set) not capturing an ACK increases, which results in duplicate packet forwarding. With a similar reasoning, O_c increases as the size of the forwarder set increases.

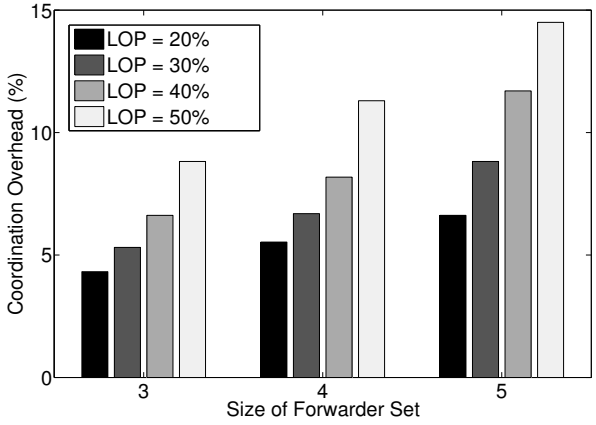


Figure 5.10: Coordination overhead for CORPL (best-effort traffic) against link outage probability (averaged over $10,000$ packets from different nodes, node density $= 3 \times 10^{-4}$ nodes per unit)

5.6 Summary and Concluding Remarks

The application of cognitive radio technology for smart grid networks is currently under active investigation in the research community. On the other hand, RPL is emerging as the de-facto routing protocol for many applications including AMI networks. A fundamental challenge in AMI networks is the reliable and low latency data delivery for different application in order to realize the vision of smart grid. Considering the promising future of cognitive smart grid networks, we propose CORPL; which is an enhanced RPL based routing protocol for cognitive radio enabled AMI networks. CORPL utilizes an opportunistic forwarding approach along with a simple overhearing based coordination scheme that not only ensures protection to PUs but also fulfils the utility requirements of the secondary network. CORPL also employs two different techniques for mitigating the performance degradation due to periodic spectrum sensing state of different nodes. Results show that CORPL improves the reliability of the network while reducing harmful interference to PUs by up to 50% as well as reducing the deadline violation probability for delay sensitive traffic. Hence, CORPL provides a viable solution for practical cognitive AMI networks.

Chapter 6

Resource Allocation in LTE-based Cellular M2M/H2H Co-existence Scenarios

6.1 Introduction

3GPP Long Term Evolution (LTE) standard uses Single Carrier Frequency Division Multiple Access (SC-FDMA) [132] as the uplink multiple access scheme. The primary benefit of SC-FDMA is its low Peak-to-Average Power Ratio (PAPR), compared to Orthogonal Frequency Division Multiple Access (OFDMA), which improves the transmit power efficiency for mobile terminals [133].

Resource allocation in OFDMA systems has been extensively studied in literature. However, existing work on resource allocation in OFDMA systems is not directly applicable to SC-FDMA systems due to the specific power and resource block (RB) allocation constraints of the latter [134], which complicate the resource allocation problem and also render the standard *Lagrangian* duality based optimization framework (as used for OFDMA systems) inapplicable.

On the other hand, Quality-of-Service (QoS) provisioning for various types of delay sensitive services is an important issue that needs to be considered in most resource allocation problems. The issue becomes particularly challenging with the introduction of Machine-to-Machine (M2M) communications (also known as Machine Type Communications (MTC)) [17] in LTE networks as the available resources are shared between M2M devices and Human-to-Human (H2H) users, having different and often conflicting QoS requirements. Conventional resource allocation algorithms for H2H typically focus on throughput maximization which may not hold for M2M scenarios as MTC mostly consists of low data rate applications. However, delay requirements for MTC can be critical

especially for applications where near real time decision making is involved such as smart grids. Therefore, simultaneous consideration of delay and data rate requirements completely characterizes the M2M/H2H co-existence scenario. It should be noted that deterministic delay guarantees cannot be provided due to the time varying nature of wireless channel.

Recently, energy efficiency has become an important objective in resource allocation due to the growing proliferation of smartphones (and other similar high-end devices) and energy-hungry applications. The requirements of energy efficiency become critical with the introduction of MTC as majority of M2M devices are battery operated and often deployed in areas where frequent human access or battery replacement is not always feasible.

In this chapter, we address the problem of energy efficient resource allocation for the uplink of LTE networks (or SC-FDMA uplink systems) under statistical QoS guarantees. To the best of our knowledge, this problem has not been investigated before in general or in M2M/H2H co-existence scenarios.

6.1.1 Related Work on Resource Allocation in the Uplink of LTE Networks

Previous works in [135] and [136] address resource allocation in SC-FDMA uplink systems through heuristic algorithms. However, the former only considers throughput maximization whereas the latter considers an average end-to-end delay requirement. Similar to [135], the authors in [137] consider sum throughput maximization, but under robust modulation and coding scheme constraints. The authors in [138] consider an M2M/H2H co-existence scenario and propose a resource allocation algorithm with the primary objective of maximizing the aggregate network utility. However, the utility for both M2M devices and H2H users is considered to be a function of data rate only. Moreover, none of the above mentioned works considers energy efficiency requirements. The authors in [139], [140] focus on resource allocation in SC-FDMA uplink systems with the objective of transmit power minimization. The authors in [141] consider a similar objective with M2M/H2H co-existence and an average delay constraint for M2M devices. However, there is no mention of energy efficiency and statistical QoS provisioning in these works.

6.1.2 Contributions and Outline

Our contributions with respect to this chapter are stated as follows.

- We consider an M2M/H2H co-existence scenario in the uplink of LTE networks and model the energy efficient resource allocation problem as the maximization of *bits-per-joule capacity* [142] under QoS requirements of different users as well as specific power and RB allocation constraints of SC-FDMA. The concept of *effective capacity* [143] is used to model the statistical delay QoS requirements of different users.
- Since standard Lagrangian duality techniques cannot be applied as discussed later, we solve the resource allocation problem using *canonical* duality theory [11], [144]. We first transform the original problem into a *Mixed Integer Programming* (MIP) problem and then convert it into a continuous space canonical dual problem that is a concave maximization problem. We identify the conditions under which the solution of canonical dual problem is identical to the primal problem.
- The canonical dual problem results in complex non-linear equations which have been efficiently solved using *Invasive Weed Optimization* [12] algorithm.
- In addition to the complexity of the SC-FDMA design problem, the consideration of statistical QoS requirements introduces a complicated expected value in the objective function as well as in the constraints, which requires the Probability Density Function (PDF) of *effective* SNR (Signal-to-Noise Ratio) [145] for SC-FDMA based systems. We derive this PDF, which can be considered as a contribution in itself.

The rest of the chapter is organized as follows. Section 6.2 describes the system model. In Section 6.3, we formulate the energy efficient resource allocation problem. Section 6.4 presents the canonical dual framework. This is followed by numerical assessment in Section 6.5. Finally, Section 6.6 summarizes and concludes the paper.

6.2 System Model

We focus on the uplink of 3GPP LTE networks and consider a single cell, multi-user M2M/H2H co-existence scenario as shown in Fig. 6.1. We assume a total of K users in the coverage area of the eNB. The conventional H2H users (UEs) are indexed by

Table 6.1: Frequently used notations

Notation	Description
R_k	Data rate of k^{th} user
$\mathcal{L}$	Set of available RBs
$\mathcal{L}_k$	Set of RBs allocated to k^{th} user
$P_{k,l}$	Transmit power of k^{th} user over l^{th} RB
$\gamma_{k,l}$	SNR of k^{th} user over l^{th} RB
$\gamma_{eff,k}$	Effective SNR of k^{th} user
θ_k	Statistical QoS exponent of k^{th} user
d_k^T	Delay bound of k^{th} user
μ	Power control policy
C_E^k	Effective capacity of k^{th} user
$\mathcal{C}_j$	Bits-per-Joule capacity
P_s	Sum power in the uplink
$\boldsymbol{x}$	RB indicator vector

the set $\mathcal{H} \triangleq \{1, \cdots, h, \cdots, H\}$. It should be noted that MTC in LTE networks can be realized in three different scenarios [138]. The first scenario is similar to a UE connected with the eNB, where the MTC devices (MTCDs) directly communicate with the eNB. We index the directly connected MTCDs by the set $\mathcal{D} \triangleq \{1, \cdots, d, \cdots, D\}$. Secondly, the MTCDs can form a capillary network (device area network) wherein communication with the eNB takes place via a newly proposed architectural enhancement to LTE networks, known as the MTC gateway (MTCG). The MTCDs connected via the MTCG are indexed by the set $\mathcal{M} \triangleq \{1, \cdots, m, \cdots, M\}$. The third scenario is of Peer-to-Peer (P2P) communication between neighboring MTCDs using uplink or downlink resources. In literature, this type of communication is also known as Device-to-Device (D2D) communication. For the sake of simplicity, P2P MTCD communication is not considered in the resource allocation problem formulation. Moreover, D2D communication has its unique signaling requirements and therefore in literature, resource allocation problems for D2D are investigated separately.

In LTE networks, the radio resources are distributed in both time and frequency domains as shown in Fig. 6.2. In the time domain, radio resources are distributed every Transmission Time Interval (TTI) which consists of two slots and has a duration of 1 ms. 20 slots or 10 TTIs constitute one LTE frame. In the frequency domain, the available bandwidth is

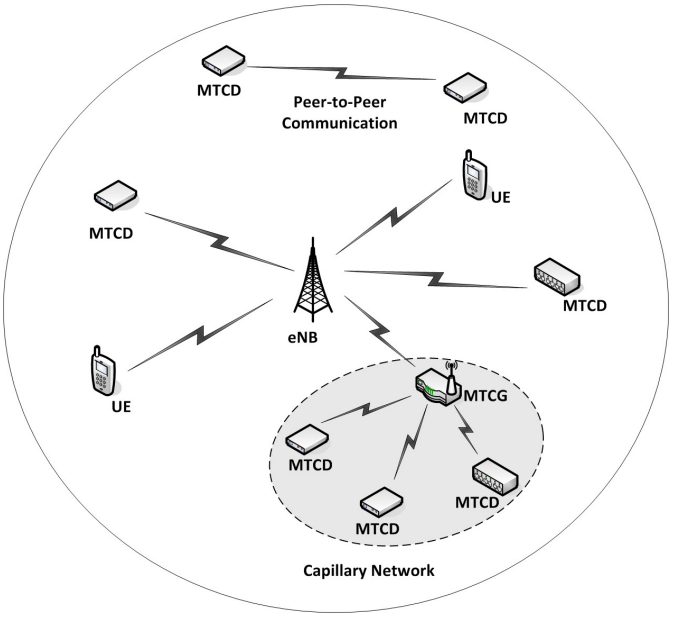


Figure 6.1: Single cell, multi-user M2M/H2H co-existence scenario

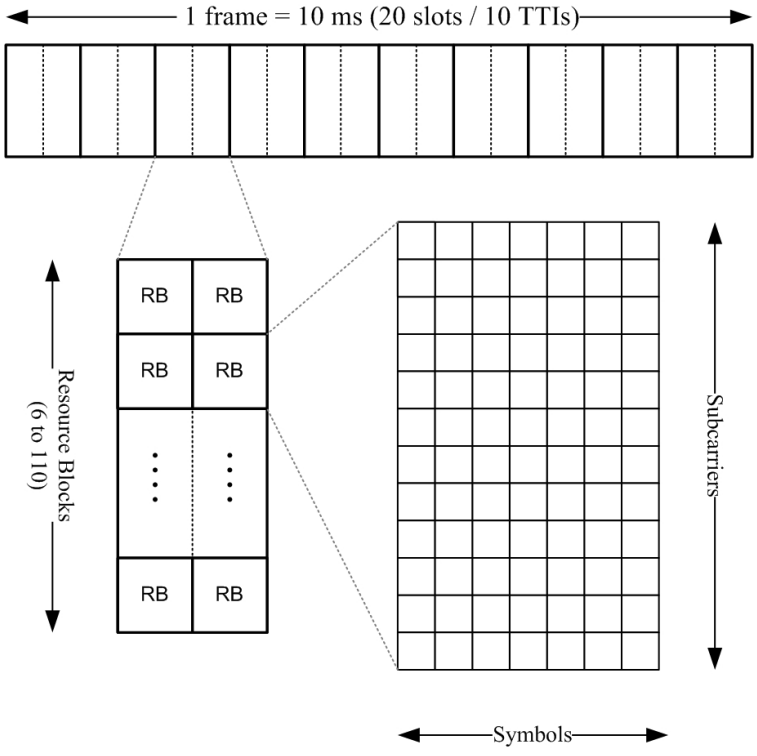


Figure 6.2: LTE frame structure

divided into a number of sub-channels each including 12 subcarriers. Each sub-channel has a bandwidth of 180 KHz and along with 7 symbols in the time domain constitutes a Resource Block (RB). Depending on the available bandwidth, the number of RBs can vary from 6 to 110.

Further, we assume that the total number of available RBs are indexed by the set $\mathcal{L} \triangleq \{1, \cdots, l, \cdots, L\}$. The channel is assumed to exhibit block fading characteristics. The coherence time of the channel is greater than the Transmission Time Interval (TTI), so it stays relatively constant during the TTI.

The use of SC-FDMA in the uplink implies certain restrictions on power and RB allocation [132]. Firstly, a single RB can only be allocated to at most one user (*exclusivity* restriction). Secondly, multiple RBs allocated to a user must be adjacent (*adjacency* restriction). Thirdly, the transmit power on all RBs allocated to a user should be equal (for retaining the low PAPR benefits).

We assume that a set of consecutive RBs, $\mathcal{L}_h$, is allocated to a UE h in the current TTI. The Signal-to-Noise Ratio (SNR) for the h^{th} UE over the l^{th} RB is given by $\gamma_{h,l} = \frac{P_{h,l}|h_{h,l}|^2}{\sigma^2}$, where $P_{h,l}$ is the transmission power of the h^{th} UE over the l^{th} RB, $h_{h,l}$ is the channel fading coefficient, and σ^2 denotes the power of the Additive White Gaussian Noise (AWGN). Using Shannon's capacity formula, the upper bound on the achievable data rate (in bits per second) for the h^{th} UE is given by

$$R_h = B \cdot L_h \log_2(1 + \mu \cdot \gamma_{eff,h}), \tag{6.1}$$

where B is the bandwidth of each RB, $L_h = \mid \mathcal{L}_h \mid$ denotes the cardinality of the set $\mathcal{L}_h$, $\gamma_{eff,h}$ is the effective SNR [145] for the h^{th} UE, and μ denotes the power control policy [146] to be discussed later. As per [147], the effective SNR for SC-FDMA symbol cannot be approximated using the same measures as used for OFDMA (e.g., Exponential Effective SNR Mapping (EESM), Mutual Information Effective SNR Mapping (MIESM) [145]) due to the fact that SC-FDM transmission spreads each data symbol over the whole bandwidth and therefore channel gain differences are averaged out over a sufficiently large bandwidth, even though each subcarrier experiences a different channel gain. Thus, the effective SNR

can be computed as the average SNR over a set of RBs and is given by

$$\gamma_{eff,h} = \frac{1}{L_h} \sum_{l \in \mathcal{L}_h} \gamma_{h,l}, \tag{6.2}$$

where $\gamma_{h,l}$ is the instantaneous SNR for the h^{th} UE over the l^{th} RB. Similarly, assuming that a set of consecutive RBs, $\mathcal{L}_d$, is allocated to the d^{th} MTCD (directly connected to the eNB) in the current TTI, the upper bound on its achievable data rate (in bits per second) is given by $R_d = B \cdot L_d \log_2(1 + \mu \cdot \gamma_{eff,d})$, where $L_d = \mid \mathcal{L}_d \mid$ denotes the cardinality of the set $\mathcal{L}_d$, and $\gamma_{eff,d}$ can be calculated in a similar way as in (6.2).

As the MTCG operates in half-duplex mode, we assume that the eNB allocates a set of consecutive RBs, $\mathcal{L}_m$, for the overall eNB-MTCG-MTCD link in the current TTI. These RBs are shared in the time domain between access and backhaul slots. Thus we have a two hop channel: MTCD-MTCG in the access slot, and MTCG-eNB in the backhaul slot. A simple fairness scheme such as Round Robin can be employed by the MTCG for the MTCDs connected to it. The expression for the upper bound on achievable data rate (in bits per second) in the access slot for the m^{th} MTCD, is given by $R_m^{(a)} = B \cdot L_m \log_2(1 + \mu \cdot \gamma_{eff,m}^{(a)})$, where $L_m = \mid \mathcal{L}_m \mid$ denotes the cardinality of the set $\mathcal{L}_m$, and $\gamma_{eff,m}^{(a)}$ is the effective SNR for the m^{th} MTCD in the access slot. Similarly, the corresponding expression for the upper bound on achievable data rate (in bits per second) in the backhaul slot is given by $R_m^{(b)} = B \cdot L_m \log_2(1 + \mu \cdot \gamma_{eff,m}^{(b)})$. The overall achievable data rate for the m^{th} MTCD can be calculated as $R_m = \frac{1}{2}\min\left(R_m^{(a)}, R_m^{(b)}\right)$.

Next, we define $\boldsymbol{\theta} = [\theta_k]_{K \times 1}$, with θ_k being the statistical QoS exponent of the k^{th} user, where $k \in \mathcal{K} = \mathcal{H} \cup \mathcal{D} \cup \mathcal{M}$. The QoS exponent θ_k characterizes the steady state delay violation probability of the k^{th} user such that $Pr\left\{d_k > d_k^T\right\} = e^{-\theta_k \varphi d_k^T}$, where d_k is the delay, d_k^T is the delay bound, and φ is determined by the arrival and service processes [148]. It should be noted that a smaller θ_k indicates a looser QoS constraint whereas a larger θ_k implies a more stringent QoS requirement.

As discussed in Appendix F, the effective capacity [143] for the k^{th} user, defined as the maximum constant arrival rate under the statistical QoS requirement specified by θ_k, is

given by

$$C_E^k(\theta_k) = -\frac{1}{\theta_k} \ln \mathbb{E}\left(e^{-\theta_k R_k}\right), \tag{6.3}$$

where $\mathbb{E}(\cdot)$ denotes the expectation operation (see Appendix G for more details) and R_k is the data rate of the k^{th} user.

The relationship between R_k, θ_k, and allocated power is given by the power control policy, μ. Conventionally, the power control policy is expressed as a function of SNR only. However in our case, it is a function of both SNR and QoS exponent (i.e., $\mu \rightarrow \mu(\theta_k, \gamma_{eff,k})$).

Lemma 6.2.1 *The optimal power control policy (which is a function of θ_k and $\gamma_{eff,k}$) [146] that maximizes the effective capacity for the k^{th} user is given by $\mu_{opt,k} = \frac{1}{\gamma_{eff,k}}\left[\left(\frac{\gamma_0}{\gamma_{eff,k}}\right)^{\frac{1}{q-1}} - 1\right]$, where $q = -\frac{\theta_k B L_k}{\ln 2}$ and γ_0 is the cutoff effective SNR.*

Proof The proof, which is obtained using the PDF of $\gamma_{eff,k}$ (derived in Appendix G), is given in Appendix H.

6.3 Problem Formulation

6.3.1 Energy Efficient Design Problem

Let P_s denote the cumulative transmit power of all the users in the uplink, given by $P_s = \sum_{k=1}^{K}\sum_{l=1}^{L} P_{k,l}$, where $P_{k,l}$ denotes the transmit power for the k^{th} user over the l^{th} RB. The power allocation for a user is subject to the condition $\sum_{l=1}^{L_k} P_{k,l} \leq P_{max}$, where P_{max} represents the maximum allowable transmit power per user and L_k denotes the cardinality of the set of RBs allocated to the k^{th} user. We formulate the energy efficient design problem in terms of overall bits-per-joule capacity for the uplink as the ratio of overall effective capacity and the sum power for all the users as follows.

$$\mathcal{C}_j(\boldsymbol{\theta}, P_s) = \frac{\sum_{k=1}^{K} C_E^k(\theta_k)}{P_s} = \frac{\sum_{k=1}^{K} -\frac{1}{\theta_k} \ln \mathbb{E}\left(e^{-\theta_k R_k}\right)}{\sum_{k=1}^{K}\sum_{l=1}^{L} P_{k,l}} \tag{6.4}$$

It is important to mention that the above definition of bits-per-joule capacity is

different from the conventional definition as it considers the QoS requirements as well. On the contrary, the conventional definition regards $\theta_k \to 0$ for all the users, indicating a system without any delay requirement. The energy efficient resource allocation problem is formulated as the maximization of bits-per-joule capacity subject to different constraints, given as follows.

$$\max_{\mathcal{L}_k, P_{k,l}} \ \mathcal{C}_j(\boldsymbol{\theta}, P_s) \tag{6.5}$$

$$\begin{aligned}
s.t \quad & (a) \quad R_k \geq {R_k}^{min} \quad \forall \quad k \in \mathcal{K} \\
& (b) \quad P_{k,l} = P_{k,n} \quad \forall \quad k \in \mathcal{K} \quad \& \quad \forall \quad l, n \in \mathcal{L} \\
& (c) \quad \mathcal{L}_k \cap \mathcal{L}_j = \varnothing \quad \forall \quad k \neq j, \quad k, j \in \mathcal{K} \\
& (d) \quad \{\min(\mathcal{L}_k), \min(\mathcal{L}_k)+1, \cdots, \max(\mathcal{L}_k)\} \cap \{\mathcal{L}_1 \\
& \qquad \cup \cdots \mathcal{L}_{k-1} \cup \mathcal{L}_{k+1} \cup \cdots \mathcal{L}_k\} = \varnothing \quad \forall \quad k \in \mathcal{K} \\
& (e) \quad \mathbb{E}\left(e^{-\theta_k R_k}\right) \leq e^{-\theta_k \mathcal{B}_k} \quad \forall \quad k \in \mathcal{K}
\end{aligned}$$

The objective of the optimization problem in (6.5) is power and RB allocation for different users in order to maximize the overall bits-per-joule capacity subject to different constraints of SC-FDMA along with satisfying the data rate and statistical delay QoS requirements. The constraint (6.5b) ensures that the transmit power on all the RBs allocated to a user is equal. The constraints (6.5c) and (6.5d) respectively reflect the exclusivity and adjacency restrictions of RB allocation in SC-FDMA. In order to guarantee a QoS constraint, θ_k, we should have $C_E^k(\theta_k) \geq \mathcal{B}_k(\theta_k)$, where $\mathcal{B}_k(\theta_k)$ is the *effective bandwidth* (dual function of effective capacity; see Appendix F for more details) of the k^{th} user. Consequently, by using the expression in (6.3), we have the constraint (6.5e)[1] [149].

It should be noted that the optimization problem (6.5) is difficult to solve due to its combinatorial nature. The cardinality of the feasible search space for K users and L RBs is given by $\sum_{i=1}^{K} \binom{K}{i} i! \binom{L-1}{i-1}$ [135], which increases exponentially. For example, for $K = 10$ and $L = 24$, the optimal solution requires a search across 5.26×10^{12} possible RB allocations,

[1]For the sake of brevity, we drop the argument of $\mathcal{B}_k(\theta_k)$ in constraint (6.5e).

which is not practical. Also note that standard Lagrangian based duality techniques, as developed for OFDMA based systems, cannot be used due to the adjacency restriction of RBs.

6.3.2 MIP Formulation

As an intermediate step towards its solution, we convert the optimization problem (6.5) into a *Mixed Integer Programming* (MIP) problem by defining a RB allocation matrix [135] which ensures the adjacency of allocated RBs for each user. The RB allocation matrix is of the order $\mid \mathcal{L} \mid \times A$, where each row corresponds to the RB index and each column corresponds to a feasible (meeting adjacency restriction) RB allocation pattern, and A denotes the total number of feasible allocation patterns such that $A = 0.5 \times (\mid \mathcal{L} \mid^2 + \mid \mathcal{L} \mid)$. The basic idea of this RB allocation matrix, for the case of 4 RBs is illustrated by (6.6). In any allocation pattern (i.e., any column), a '1' is placed when the RB is allocated to a user, otherwise a '0' is placed.

$$\boldsymbol{M}^k = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \tag{6.6}$$

We define a RB indicator vector $\boldsymbol{x} \triangleq [\boldsymbol{x}_k]_{K\times 1}$, where $\boldsymbol{x}_k = [x_{k,a}]_{A\times 1}$, such that each entry $x_{k,a} \in \{0, 1\}$ indicates whether the RB allocation pattern a is allocated to the k^{th} user or not. Thus, we rewrite (6.5) as a MIP problem as follows.

$$(\mathcal{P}_{min}) : \min_{\boldsymbol{x}} \left\{ \mathcal{C}_j(\boldsymbol{\theta}, \boldsymbol{x}) = \frac{\sum_{k=1}^{K} \frac{1}{\theta_k} \ln \mathbb{E}\left(e^{-\theta_k R_k}\right)}{\sum_{k=1}^{K} \sum_{a=1}^{A} x_{k,a} P_{k,a}} \right\} \tag{6.7}$$

$$s.t \quad (a) \quad \mathbb{E}\left(e^{-\theta_k R_k}\right) \leq e^{-\theta_k \mathcal{B}_k} \quad \forall \quad k \in \mathcal{K}$$

$$(b) \quad x_{k,a}(x_{k,a} - 1) = 0 \qquad \forall \quad k \in \mathcal{K} \;\; \& \;\; \forall \quad a$$

$$(c) \quad \sum_{k=1}^{K} \sum_{a=1}^{A} x_{k,a} M_{l,a}^{k} = 1 \quad \forall \quad l \in \mathcal{L}$$

$$(d) \quad \sum_{a=1}^{A} x_{k,a} = 1 \qquad \forall \quad k \in \mathcal{K}$$

The optimization problem (6.7) is an MIP equivalent of the optimization problem (6.5). The constraint (6.7a) is the same as constraint (6.5e). The constraint (6.7b) is a pure binary constraint that ensures that $x_{k,a} \in \{0, 1\}$. The constraint (6.7c), where $M_{l,a}^{k}$ denotes the l^{th} row and a^{th} column of the matrix $\boldsymbol{M}^{k}$, ensures the exclusivity of allocated RBs. The constraint (6.7d) ensures that at most one allocation pattern is chosen for each user. The constraints (6.5a) and (6.5b) are implicitly accommodated in $P_{k,a}$ which denotes the optimal transmit power for the k^{th} user when the a^{th} allocation pattern is used and is obtained by setting $R_k = R_k^{min}$ (employing the optimal power control policy, $\mu_{opt,k}$ given by *Lemma 1*), and $P_{k,l} = \frac{P_{k,a}}{L_{k,a}}$, where $L_{k,a}$ denotes the cardinality of the set of RBs ($\mathcal{L}_{k,a}$) allocated to user k, when the a^{th} allocation pattern is used.

Although the optimization problem (6.7) is more tractable and appears to be simpler than (6.5), the solution is still exponentially complex.

For comparison, we also formulate a spectral efficient design problem[2] where the objective is to maximize the overall effective capacity under the same constraints as in (6.7). It should be noted that the spectral efficient design always allocates the maximum allowable transmit power per user, whereas the energy efficient design dynamically adjusts the level of transmit power.

[2]The spectral efficient design problem is given by $\max \left\{ \mathcal{C}_s(\boldsymbol{\theta}) = \sum_{k=1}^{K} C_E^k(\theta_k) = \sum_{k=1}^{K} -\frac{1}{\theta_k} \ln \mathbb{E}\left(e^{-\theta_k R_k}\right) \right\}$ subject to (6.7a) – (6.7d).

6.4 Canonical Dual Framework

6.4.1 Canonical Duality Theory

Canonical Duality Theory (CDT) [11], [144] was developed from non-convex analysis and mechanics during the last decade. The theory comprises of *canonical dual transformation* methodolgy, a *complementary-dual principle*, and an associated *triality* theory. The canonical dual transformation is used to convert the non-smooth problem into a smooth canonical dual problem with zero duality gap, the complementary-dual principle provides the relationship between the primal and the dual problem, and the triality theory, whose components comprise a saddle min-max duality, helps in identifying local and global extrema.

6.4.2 Dual Problem Formulation

Next, we convert our MIP problem (6.7) into a continuous space canonical dual problem using CDT. Since the canonical dual problem is solved in continuous space, we provide the conditions under which the solution of canonical dual problem is identical to the solution of respective primal problem. A generic framework for solving 0-1 quadratic problems using CDT is given in [150] which has been extended in [151] and [152] for solving different resource allocation problems. However, our problem is much more complex because of a difficult objective function, additional constraints, and its MIP nature. A framework for solving general MIP problems using CDT is given in [153], which will be extended to solve (6.7).

The feasible space for the primal problem in (6.7) is defined by $\mathcal{X}_p = \{(\boldsymbol{\theta}, \boldsymbol{x}) \in \mathbb{R}^K \times \{0,1\}^{KA\times 1}\}$. We temporarily relax the equality constraints in (6.7) to inequalities, solve the problem in the continuous space, and provide the conditions under which the solution of canonical dual problem is identical to the solution of respective primal problem.

As a key step towards canonical dual formulation, we introduce a non-linear transformation (geometrical mapping): $\mathbf{y} = \wedge(\boldsymbol{\theta}, \boldsymbol{x}) = (\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\sigma}) \in \mathcal{Y}_g$, which is a vector valued mapping where $\mathcal{Y}_g$ denotes the feasible space for $\mathbf{y}$, and

$$\begin{cases} \boldsymbol{\delta} &= \left[\mathbb{E}\left(e^{-\theta_k R_k}\right) - e^{-\theta_k \mathcal{B}_k}\right]_{K\times 1} \\ \boldsymbol{\beta} &= \left[x_{k,a}(x_{k,a}-1)\right]_{KA\times 1} \\ \boldsymbol{\tau} &= \left[\sum_{k=1}^{K}\sum_{a=1}^{A} x_{k,a} M_{l,a}^{k} - 1\right]_{L\times 1} \\ \boldsymbol{\sigma} &= \left[\sum_{a=1}^{A} x_{k,a} - 1\right]_{K\times 1} \end{cases} \tag{6.8}$$

Thus, the feasible space for $\mathbf{y}$ is given by $\mathcal{Y}_g = \mathbb{R}^K \times \mathbb{R}^{KA} \times \mathbb{R}^L \times \mathbb{R}^K \mid \boldsymbol{\delta} \leq 0, \boldsymbol{\beta} \leq 0, \boldsymbol{\tau} \leq 0, \boldsymbol{\sigma} \leq 0$.

Next, we define the indicator function (canonical function) [153] of the constraints as follows.

$$V(\mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{y} \leq 0 \\ +\infty & \text{otherwise} \end{cases} \tag{6.9}$$

Using the indicator function in (6.9), we rewrite the primal problem (6.7) in the canonical form as follows.

$$\min\left\{V\left(\wedge(\boldsymbol{\theta}, \boldsymbol{x})\right) + \mathcal{C}_j(\boldsymbol{\theta}, \boldsymbol{x})\right\} \tag{6.10}$$

Let, $\mathbf{y}^* = (\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*)$ be the vector of dual variables associated with the corresponding restrictions $\mathbf{y} \leq 0$. The feasible space for the existence of dual variables is given by $\mathcal{Y}_d = \mathbb{R}^K \times \mathbb{R}^{KA} \times \mathbb{R}^L \times \mathbb{R}^K \mid \boldsymbol{\delta}^* \geq 0, \boldsymbol{\beta}^* \geq 0, \boldsymbol{\tau}^* \geq 0, \boldsymbol{\sigma}^* \geq 0$.

According to the Fechnel transformation, the canonical sup-conjugate function associated with $V(\mathbf{y})$ is defined as follows.

$$\begin{aligned} V^*(\mathbf{y}^*) &= \sup\left\{\langle \mathbf{y}, \mathbf{y}^* \rangle - V(\mathbf{y}) \mid \mathbf{y} \in \mathcal{Y}_g, \mathbf{y}^* \in \mathcal{Y}_d\right\} \\ &= \sup_{\mathbf{y}^*}\left\{\boldsymbol{\delta}^T\boldsymbol{\delta}^* + \boldsymbol{\beta}^T\boldsymbol{\beta}^* + \boldsymbol{\tau}^T\boldsymbol{\tau}^* + \boldsymbol{\sigma}^T\boldsymbol{\sigma}^* - V(\mathbf{y})\right\} \\ &= \begin{cases} 0 & \text{if } \boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^* \geq 0 \\ +\infty & \text{otherwise} \end{cases} \end{aligned} \tag{6.11}$$

Since $V(\mathbf{y})$ is a proper closed convex function over $\mathcal{Y}_g$, the following canonical duality relations exist:

$$\mathbf{y}^* \in \partial V(\mathbf{y}) \quad \Leftrightarrow \quad \mathbf{y} \in \partial V^*(\mathbf{y}^*) \quad \Leftrightarrow \quad V(\mathbf{y}) + V^*(\mathbf{y}^*) = \mathbf{y}^T\mathbf{y}^*.$$

Using the definition of sub-differential, it can be easily verified that if $\mathbf{y}^* > 0$, then the condition $\mathbf{y}^T\mathbf{y}^* = 0$ leads to $\mathbf{y} = 0$, and consequently $(\boldsymbol{\theta}, \boldsymbol{x}) \in \mathcal{X}_p$. Therefore, the dual feasible space for the primal problem in (6.7) is an open positive cone defined by $\mathcal{X}_p^\sharp = \{\mathbf{y}^* \in \mathcal{Y}_d \mid \mathbf{y}^* > 0\}$.

By using the Fechnel-Young equality, $V(\wedge(\boldsymbol{\theta}, \boldsymbol{x})) = \wedge(\boldsymbol{\theta}, \boldsymbol{x})^T\mathbf{y}^* - V^*(\mathbf{y}^*)$, the *total complementarity function* [11] is defined as follows.

$$\Xi(\boldsymbol{\theta}, \boldsymbol{x}, \mathbf{y}^*) = \wedge(\boldsymbol{\theta}, \boldsymbol{x})^T\mathbf{y}^* - V^*(\mathbf{y}^*) + \mathcal{C}_j(\boldsymbol{\theta}, \boldsymbol{x}) \tag{6.12}$$

Using the definitions of $\wedge(\boldsymbol{\theta}, \boldsymbol{x})$, $V^*(\mathbf{y}^*)$, and $\mathcal{C}_j(\boldsymbol{\theta}, \boldsymbol{x})$, we can express $\Xi(\boldsymbol{\theta}, \boldsymbol{x}, \mathbf{y}^*) = \Xi(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*)$ as follows.

$$\begin{aligned}\Xi(\boldsymbol{\theta}, \boldsymbol{x}, \mathbf{y}^*) = &\sum_{k=1}^{K}\sum_{a=1}^{A} x_{k,a}\left(\sigma_k^* - \beta_{k,a}^* + \sum_{l=1}^{L}\tau_l^* M_{l,a}^k\right) + \frac{\sum_{k=1}^{K}\frac{1}{\theta_k}\ln\mathbb{E}\left(e^{-\theta_k R_k}\right)}{\sum_{k=1}^{K}\sum_{a=1}^{A} x_{k,a}P_{k,a}} \\ &- \sum_{k=1}^{K}\sigma_k^* - \sum_{l=1}^{L}\tau_l{}^* + \sum_{k=1}^{K}\delta_k^*\mathbb{E}\left(e^{-\theta_k R_k}\right) - \sum_{k=1}^{K}\delta_k^* e^{-\theta_k \mathcal{B}_k} + \sum_{k=1}^{K}\sum_{a=1}^{A}\beta_{k,a}^* x_{k,a}^2\end{aligned} \tag{6.13}$$

Next, we define the *canonical dual function* [11], [150] using the canonical dual variables (defined in the dual space) as follows.

$$\Omega(\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*) = \mathrm{sta}\left\{\Xi(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*)\right\}, \tag{6.14}$$

where $\mathrm{sta}(\cdot)$ denotes finding the stationary point of the function. We are primarily interested in the allocation vector $\boldsymbol{x}_k$ for a user k. Therefore, for a given QoS constraint, θ_k, the stationary point of $\Xi(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*)$ occurs at

$$x_{k,a}(\theta_k, \mathbf{y}^*) = \frac{\ln\mathbb{E}\left(e^{-\theta_k R_k}\right)}{2\theta_k P_{k,a}\beta_{k,a}^*} - 0.5\frac{w}{\beta_{k,a}^*}, \tag{6.15}$$

where $w = \left(\sum_{l=1}^{L}\tau_l^* M_{l,a}^k + \sigma_k^* - \beta_{k,a}^*\right)$ and the stationary point is obtained by $\nabla_{\boldsymbol{x}}\Xi(\boldsymbol{\theta}, \boldsymbol{x}, \mathbf{y}^*) = 0$. Using (6.14) and (6.15), we obtain the dual function, which is given by

$$\Omega(\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*) = \sum_{k=1}^{K} \delta_k^* \mathbb{E}\left(e^{-\theta_k R_k}\right) - \sum_{k=1}^{K} \delta_k^* e^{-\theta_k \mathcal{B}_k} - \sum_{k=1}^{K} \sigma_k^* - \sum_{l=1}^{L} {\tau_l}^* + \frac{\sum_{k=1}^{K} \frac{1}{\theta_k} \ln \mathbb{E}\left(e^{-\theta_k R_k}\right)}{\sum_{k=1}^{K} \sum_{a=1}^{A} \frac{\Psi}{2\beta_{k,a}^*}} + \sum_{k=1}^{K} \sum_{a=1}^{A} \frac{\Psi}{2P_{k,a}\beta_{k,a}^*} \Big(\frac{\Psi}{2P_{k,a}} + w \Big) \tag{6.16}$$

where $\Psi = \theta_k^{-1} \ln \mathbb{E}\left(e^{-\theta_k R_k}\right) - P_{k,a} w$.

The dual function is a concave function on $\mathcal{X}_p^\sharp$. The canonical dual problem associated with (6.7) can be formulated as follows.

$$\min\left\{\mathcal{C}_j(\boldsymbol{\theta}, \boldsymbol{x}) \mid \mathcal{X}_p\right\} = \max\left\{\Omega(\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*) \mid \mathcal{X}_p^\sharp\right\} \tag{6.17}$$

Theorem 6.4.1 *If* $\widetilde{\mathbf{y}}^* = (\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*) \in \mathcal{X}_p^\sharp$ *denote the KKT point of the dual function such that* $\widetilde{\boldsymbol{x}}$ *denotes the KKT point of the primal problem (for a given* $\boldsymbol{\theta}$*) and* $\mathcal{C}_j(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}) = \Omega(\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*)$*, then there exists a perfect duality relationship between the primal problem in* (6.7) *and its canonical dual problem.*

Proof The proof is given in Appendix I.

Theorem 6.4.1 (also known as complementary dual principle) shows that the MIP in (6.7) is converted into a continuous space canonical dual problem which is perfectly dual to it. Moreover, the KKT point of the dual problem provides the KKT point of the primal problem. The equivalent canonical dual problem for spectral efficient design can be obtained in a similar way as described for (6.7).

Theorem 6.4.2 *If* $\widetilde{\mathbf{y}}^* = (\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*) \in \mathcal{X}_p^\sharp$*, then* $\widetilde{\boldsymbol{x}}$ *is a global minimizer of* $\mathcal{C}_j(\boldsymbol{\theta}, \boldsymbol{x})$ *over* $\mathcal{X}_p$ *(for a given* $\boldsymbol{\theta}$*) and* $\widetilde{\mathbf{y}}^*$ *is a global maximizer of* $\Omega(\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*)$ *over* $\mathcal{X}_p^\sharp$*. Therefore,* $\mathcal{C}_j(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}) = \min\left\{\mathcal{C}_j(\boldsymbol{\theta}, \boldsymbol{x}) \mid \mathcal{X}_p\right\} = \max\left\{\Omega(\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*) \mid \mathcal{X}_p^\sharp\right\} = \Omega(\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*)$.

Proof The proof is given in Appendix J.

Theorem 6.4.2 provides the global optimality conditions, i.e., the solution of the canonical dual problem provides optimal solution to the primal problem, if the given global optimality conditions are met. Since the dual problem is a concave maximization problem over $\mathcal{X}_p^\sharp$, solving the KKT conditions associated with the dual function in (6.16) is necessary and sufficient for global optimality.

The KKT conditions of the dual function in (6.16) are given by $\partial\Omega/\partial\delta_k^* = 0$, $\partial\Omega/\partial\beta_{k,a}^* = 0$, $\partial\Omega/\partial\tau_l^* = 0$, and $\partial\Omega/\partial\sigma_k^* = 0$, where the respective partial derivatives are given by (6.18), (6.19), (6.20), and (6.21).

$$\frac{\partial\Omega}{\partial\delta_k^*} = \mathbb{E}\left(e^{-\theta_k R_k}\right) - e^{-\theta_k \mathcal{B}_k} \tag{6.18}$$

$$\begin{aligned}\frac{\partial\Omega}{\partial\beta_{k,a}^*} = \frac{\sum_{k=1}^{K} -\frac{1}{\theta_k}\ln\mathbb{E}\left(e^{-\theta_k R_k}\right)}{\sum_{k=1}^{K}\sum_{a=1}^{A}\left(P_{k,a} + \frac{\Psi}{2(\beta_{k,a}^*)^2}\right)} - \sum_{k=1}^{K}\sum_{a=1}^{A}\left\{\frac{3\Psi}{4P_{k,a}\beta_{k,a}^*}\right\} \\ - \sum_{k=1}^{K}\sum_{a=1}^{A}\left\{\left(\frac{\Psi}{2P_{k,a}} + w\right)\left(\frac{1}{2\beta_{k,a}^*} + \frac{1}{2P_{k,a}\beta_{k,a}^*}\right)\right\}\end{aligned} \tag{6.19}$$

$$\frac{\partial\Omega}{\partial\tau_l^*} = \sum_{k=1}^{K}\sum_{a=1}^{A}\left\{\frac{\Psi\sum_{l=1}^{L} M_{l,a}^k}{4P_{k,a}\beta_{k,a}^*} - \left(\frac{\Psi}{2P_{k,a}} + w\right)\left(\frac{\sum_{l=a}^{L} M_{l,a}^k}{2\beta_{k,a}^*}\right)\right\} - L - \frac{\sum_{k=1}^{K}\frac{1}{\theta_k}\ln\mathbb{E}\left(e^{-\theta_k R_k}\right)}{\sum_{k=1}^{K}\sum_{a=1}^{A}\left(\frac{P_{k,a}\sum_{l=1}^{L} M_{l,a}^k}{2\beta_{k,a}^*}\right)} \tag{6.20}$$

$$\frac{\partial\Omega}{\partial\sigma_k^*} = \sum_{k=1}^{K}\sum_{a=1}^{A}\left\{\frac{\Psi}{4P_{k,a}\beta_{k,a}^*} - \left(\frac{\Psi}{2P_{k,a}} + w\right)\left(\frac{1}{2\beta_{k,a}^*}\right)\right\} - K - \frac{\sum_{k=1}^{K}\frac{1}{\theta_k}\ln\mathbb{E}\left(e^{-\theta_k R_k}\right)}{\sum_{k=1}^{K}\sum_{a=1}^{A}\left(\frac{P_{k,a}}{2\beta_{k,a}^*}\right)} \tag{6.21}$$

6.4.3 Invasive Weed Optimization Algorithm

It should be noted that the KKT conditions associated with the dual function result in non-linear equations. In literature, traditional gradient based approaches (e.g., Secant, Muller's, Newton's, etc.) exist for solving non-linear equations. However, these algorithms show many defects like sensitivity to choice of initial values, oscillatory behavior, etc. Moreover, the complexity associated with the differentiation of the KKT conditions complicates the application of gradient based techniques. Apart from this, the calculation of step size is not easy.

We use an Invasive Weed Optimization (IWO) [12] algorithm for solving the complex non-linear equations associated with the KKT conditions. IWO is an evolutionary optimization algorithm inspired by the invasive and robust nature of weeds in growth and colonizing. It has been shown to perform better than traditional algorithms as well as other evolutionary algorithms in terms of convergence. It also has the desirable properties

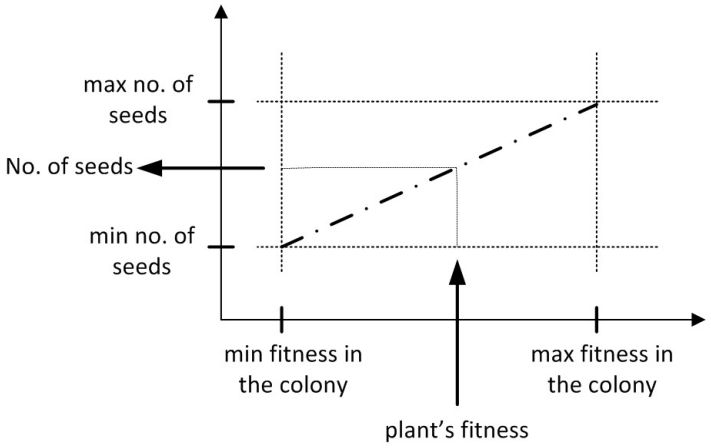


Figure 6.3: Seed production procedure in weed colony

of dealing with complex and non-differentiable objective functions and does not show the defects mentioned above.

A detailed discussion on IWO is out of scope of this chapter. Interested reader is referred to [12], [154]. We briefly describe the algorithm in terms of its key steps as follows.

- *Initialization:* In this step, a finite number of seeds are randomly dispersed over the search space as initial solutions.
- *Reproduction:* In the reproduction phase, every seed is allowed to grow to a flowering plant and is allowed to produce seeds based on its own fitness and the colony's best (f_{best}) and worst (f_{worst}) fitness as shown in Fig. 6.3.
- *Spatial Dispersion:* In this step, the produced seeds are randomly distributed in the search space following a normal distribution with zero mean and standard deviation reducing from an initial value ($\sigma_{initial}$) to a final value (σ_{final}) according to a non-linear equation given by

$$\sigma_{iter} = \left(\frac{iter_{max} - iter}{iter_{max}}\right)^{g} (\sigma_{initial} - \sigma_{final}) + \sigma_{final}, \tag{6.22}$$

  where $iter_{max}$ is the maximum number of iterations and g is the non-linear modulation index.
- *Competitive Exclusion:* Lastly, a competitive mechanism is implemented for eliminating undesirable plants with poor fitness and allowing fitter plants to produce

Table 6.2: Simulation Parameters

Parameter	Value
Cell radius	500 m
Path loss (eNB-to-UE/MTCD)	$128.1 + 37.6\log_{10}(r)$, r in km [155]
Path loss (MTCD-MTCD)	NLOS: $48.9 + 40\log_{10}(r), r \geq 0.3$, LOS: $38.5 + 20\log_{10}(r), r < 0.3$, r in m
Standard deviation of Shadowing	8 dB (90 % cell edge coverage)
Power spectral density of noise	-174 dBm/Hz
Noise figure	5 dB
QoS exponent	varied

more seeds. The process continues until $iter_{max}$ is reached or the minimum fitness threshold (ϱ) is achieved.

6.5 Numerical and Simulation Results

Our simulation model is based on the uplink of a single cell, multi user 3GPP LTE system. The system bandwidth is 10 MHz; therefore, 50 usable RBs are available per TTI, each having a bandwidth of 180 kHz. The channel model accounts for small scale Rayleigh fading, large scale path loss, and shadowing (log-normally distributed). Other simulation parameters are given in Table 6.2.

We consider three different M2M/H2H co-existence scenarios in our simulations. In *Scenario 1*, we consider 20 uniformly distributed users (10 H2H UEs and 10 directly connected MTCDs) in the coverage area with a minimum distance of 50m from the eNB. The data rate requirements for H2H users are randomly varying from 64 kbps – 576 kbps as multiples of 64 kbps, whereas in case of MTCDs, 1 RB is assumed to be sufficient to fulfill the data rate requirements. We evaluate the overall energy efficiency for both spectral efficient and energy efficient designs. Note that both spectral and energy efficient designs are based on canonical dual framework. We also compare the performance with the classical Round Robin (RR) and Best Channel Quality Indicator (BCQI) algorithms. In the RR algorithm, an equal number of contiguous RBs are allocated to each user in turn. The BCQI is a channel aware algorithm such that the user with the highest SNR is given priority.

Table 6.3: IWO numerical parameter values (per dual variable)

Parameter	**Value**
Size of initial population (Q)	20
Minimum fitness threshold (ϱ)	10^{-8}
Maximum no. of iterations ($iter_{max}$)	1000
Maximum no. of Plants (Q_{max})	10
Maximum no. of seeds (S_{max})	5
Minimum no. of seeds (S_{min})	0
Non-linear modulation index	2.5
Initial standard deviation ($\sigma_{initial}$)	10
Final standard deviation (σ_{final})	0.01

In *Scenario 2*, we consider the effect of peer-to-peer (P2P) MTCD communication i.e., MTCD – MTCD communication using uplink cellular resources. In addition to the 20 users in *Scenario 1*, we consider 10 MTCD pairs randomly distributed in the coverage area of the eNB.

In *Scenario 3*, we consider 20 users as in *Scenario 1*. However, instead of 10 directly connected MTCDs, we consider 10 MTCDs forming a capillary network and connected to the eNB via MTCG. To ensure reliable MTC, orthogonal channel allocation has been proposed for the capillary networks [138]. Therefore, we partition the set of available RBs between the capillary network and H2H users and evaluate the performance of this scheme. Note that the data rate requirements for H2H users and MTCDs in *Scenarios 2* and *3* are same as in *Scenario 1*.

In each of the above mentioned simulation scenarios, we solve the KKT conditions for each dual variable associated with the dual problem using the IWO algorithm (implemented in MATLAB) and compute the allocation vector $\boldsymbol{x}_k$ using (6.15). A pseudocode for the resource allocation algorithm is given as Algorithm 2, whereas the simulation parameters for IWO algorithm are summarized in Table 6.3.

We conducted a range of simulations and found that the choice of number of plants and the standard deviation affects the rate of convergence. However, the choice of initial search area has little effect on the final solution and the algorithm converges towards the optimal point. Fig. 6.4 shows the convergence of IWO algorithm for one of the KKT conditions;

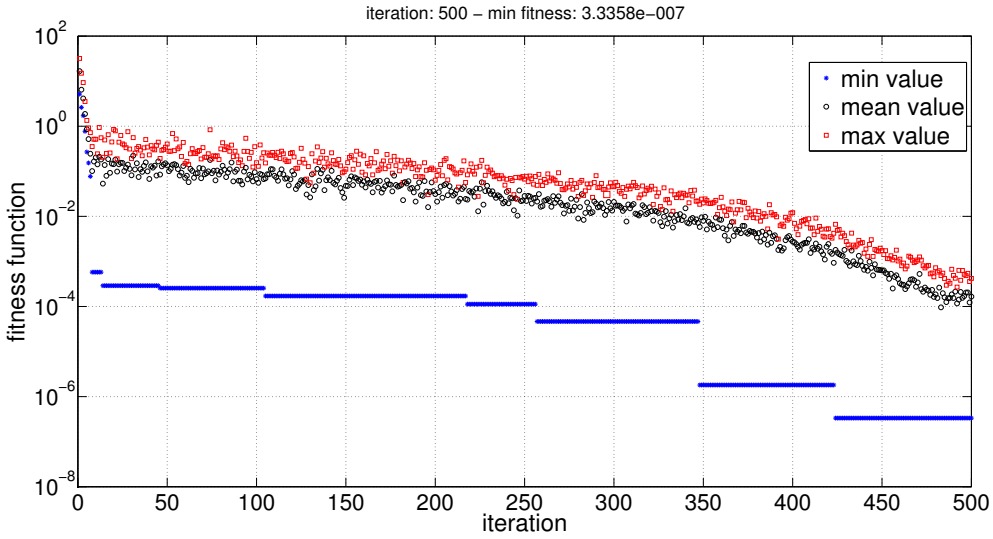


Figure 6.4: Convergence of IWO algorithm over 500 iterations for the KKT condition $\partial\Omega/\partial\sigma_k^* = 0$

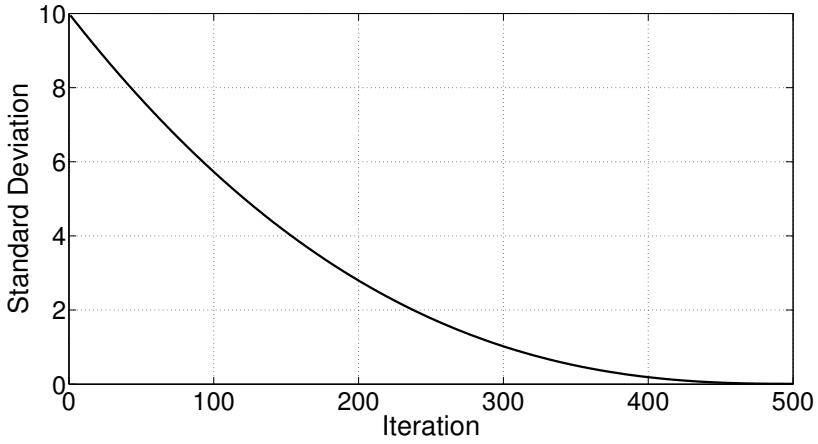


Figure 6.5: Standard deviation of spatial dispersion ($iter_{max} = 500$) for the KKT condition $\partial\Omega/\partial\sigma_k^* = 0$

Algorithm 2: Resource Allocation based on IWO

initialize $\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^* \quad \forall\, k \in \mathcal{K}, \quad iter = 0;$
$\forall \quad \partial\Omega/\partial\boldsymbol{\nu}^*, \text{where} \quad \boldsymbol{\nu}^* \in (\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*)$
create initial population of Q individuals (weeds) randomly dispersed over the search space: $\mathcal{W} = \{W_1, W_2, ..., W_Q\}$;
while $\mid \boldsymbol{\nu}^* \mid > \varrho \quad or \quad iter = iter_{max}$ **do**
  evaluate the fitness of each individual i.e., calculate $f(W_n) \quad \forall\, n \in \mathcal{W}$;
  sort $\mathcal{W}$ in ascending order according to $f(W_n)$;
  select first Q_p individuals of $\mathcal{W}$ to create the set $\mathcal{W}_p$;
  (*Reproduction*)
  $\forall\, W_j,\ j = 1, \cdots, Q_p$
  generate $S_j = \frac{f(W_j) - f_{worst}}{f_{best} - f_{worst}} \times (S_{max} - S_{min}) + S_{max}$ seeds;
  create the population of newly generated seeds, $\mathcal{W}_s = \{W_s\}$;
  (*Spatial Dispersion*)
  for $i = 1 : \mid \mathcal{W}_s \mid$ **do**
    ${W_s}^i \leftarrow {W_s}^i + \phi^i$, where $\phi^i \sim N(0, \sigma_{iter})$
  end
  (*Competitive Exclusion*)
  create the population of parents and seeds, $\mathcal{W}^* = \mathcal{W} \cup \mathcal{W}_s$;
  sort $\mathcal{W}^*$ in ascending order according to fitness;
  select first Q_{max} individuals of $\mathcal{W}^*$ and create $\mathcal{W}$;
  $iter \leftarrow iter + 1$
end
select the best fitted individuals $\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*$, and $\boldsymbol{\sigma}^*$;
if $\mathbf{y}^* = (\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*) \in \mathcal{X}_p^\sharp$ **then**
  calculate $\boldsymbol{x_k}$ using (6.15);
end
else
  $infeasible\_alloc \leftarrow infeasible\_alloc + 1;$
end

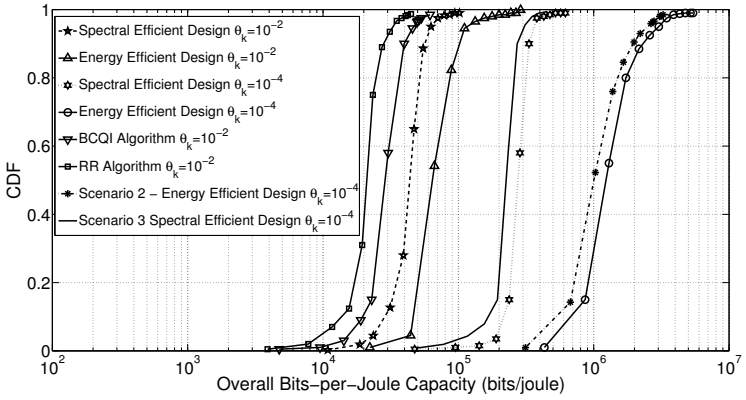
$\partial\Omega/\partial\sigma_k^* = 0$. The x-axis shows the number of iterations whereas the y-axis shows the value of fitness function, which is $\partial\Omega/\partial\sigma_k^*$ in our case. Over 500 iterations, the value of fitness function is 3.33×10^{-7}. The variation of standard deviation of spatial dispersion from $\sigma_{initial}$ to σ_{final} is shown in Fig. 6.5.

Fig. 6.6a evaluates the overall energy efficiency (in terms of bits-per-joule capacity) for different scenarios. First we discuss *Scenario 1*. We note that the energy efficient design significantly improves (e.g., for $\theta_k = 10^{-2}$ it performs 45% better in 95% of the times) the overall energy efficiency compared to the spectral efficient design as the latter always operates at the maximum allowable transmit power whereas the former adjusts the transmit power dynamically. The energy efficiency decreases as the QoS exponent, θ_k, increases. This is because a larger θ_k requires more power per RB in order to meet a more stringent delay QoS constraint[3]. We also note that both the energy efficient and spectral efficient designs outperform the RR and BCQI algorithms. Since the RR algorithm is not channel aware, its performance is worse than that of the BCQI algorithm. In *Scenario 2*, we observe a degradation in the overall energy efficiency compared to the performance without considering P2P MTCD communication (for $\theta_k = 10^{-4}$). This is due to the fact that P2P MTCD communication reuses the set of available RBs (used to serve other users) resulting in intra-cell interference and consequently degrading the overall energy efficiency. In *Scenario 3*, we note that due to RB set partitioning, the overall energy efficiency decreases (for $\theta_k = 10^{-4}$). This is because orthogonal allocation for MTCDs and H2H users results in a lack of sufficient number of RBs to fully meet the service requirements of all UEs. It should be noted that RB set partitioning approach is also spectrally inefficient.

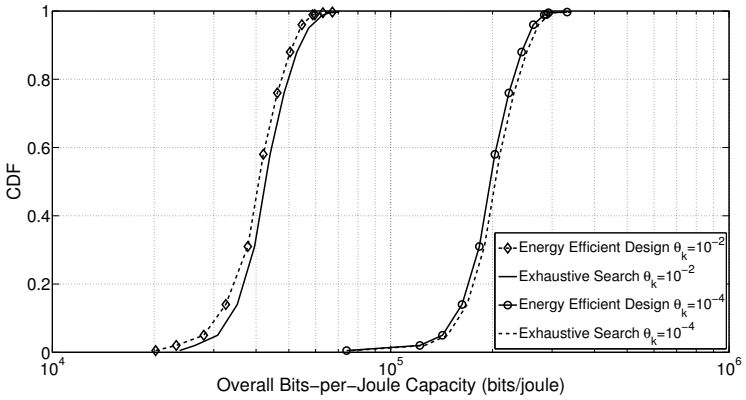
It is important to have a comparison with the optimal solution of the optimization problem (6.7). We employ exhaustive search technique to obtain the optimal solution. As discussed in Section 6.3, the cardinality of the search space increases exponentially with the number of users and RBs. Therefore, for a rather practical number of users and RBs considered in the scenarios discussed above, the exhaustive search is clearly not feasible[4]. Thus, we consider a limited scenario wherein the system bandwidth is 3 MHz (15 available

[3]A detailed discussion on power and delay trade-off in wireless networks can be found in [149].

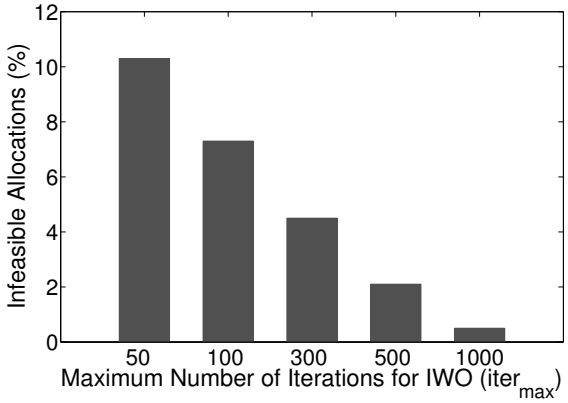
[4]Optimal solution for 20 users and 50 RBs requires a search across 8.34×10^{31} possible RB allocations, which is impractical.



(a)



(b)



(c)

Figure 6.6: (a) CDF of the overall bits-per-joule capacity (generated over 100 iterations, peak power constraint per user is assumed to be 200mW), (b) Comparison between canonical dual design and exhaustive search (CDF of overall bits-per-joule capacity generated over 200 iterations), (c) Average infeasible allocations (over 100 independent runs) against maximum number of iterations for the IWO algorithm.

RBs) and 5 uniformly distributed UEs exist with similar rate requirements as in *Scenario 1*. We compare the energy efficiency of the energy efficient design and the optimal solution. As seen from the results in Fig. 6.6b, the canonical dual design performs very close to the optimal design.

The difference arises due to infeasible allocations which result when $\mathbf{y}^* = (\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*, \boldsymbol{\sigma}^*) \notin \mathcal{X}_p^\sharp$. Our analysis shows that the infeasible allocations are mainly dependent on the maximum number of iterations ($iter_{max}$). As shown in Fig. 6.6c, the number of infeasible allocations reduces as $iter_{max}$ increases. This is because increasing the number of iterations ensures that the IWO algorithm converges to achieve the minimum fitness value ($\varrho \to 0$). In our numerical assessment of energy efficiency, we ignore the infeasible allocations. However, in practice a better approach is to compare the performance with a sub-optimal greedy algorithm and select the best solution under such a scenario. It should be noted that developing a greedy algorithm specifically designed to solve (6.7) is a separate problem by itself and rather complicated, especially under consideration of statistical QoS guarantees, and therefore out of the scope of this chapter.

Fig. 6.7 evaluates the delay violation probability for a user k as a function of the delay bound (d_k^T) for different θ_k. Recall that the delay violation probability is given by $Pr\left\{d_k > d_k^T\right\} = e^{-\theta_k \varphi d_k^T}$. Therefore, the delay violation probability follows an exponential decrease with d_k^T. It should be noted that θ_k indicates the decaying rate of QoS violation probability. A smaller θ_k corresponds to a slower decaying rate i.e., for the same d_k^T, as θ_k decreases, the delay violation probability increases, which implies that the system can only provide a looser QoS guarantee. On the other hand, a larger θ_k leads to a faster decaying rate, which means that a more stringent QoS requirement can be guaranteed. The parameter φ in delay violation probability is obtained using the effective capacity and effective bandwidth relationship from [148], which states that φ gives the rate at which the effective capacity and effective bandwidth curves intersect i.e., $C_E^k(\theta_k) = \mathcal{B}_k(\theta_k)$.

Similarly, Fig. 6.8 evaluates the delay violation probability for a user k as a function of delay bound (d_k^T) for different $\gamma_{eff,k}$. For a fixed d_k^T, the delay violation probability decreases as $\gamma_{eff,k}$ increases, which implies that the system can provide a more stringent QoS guarantee by allocating more power per RB.

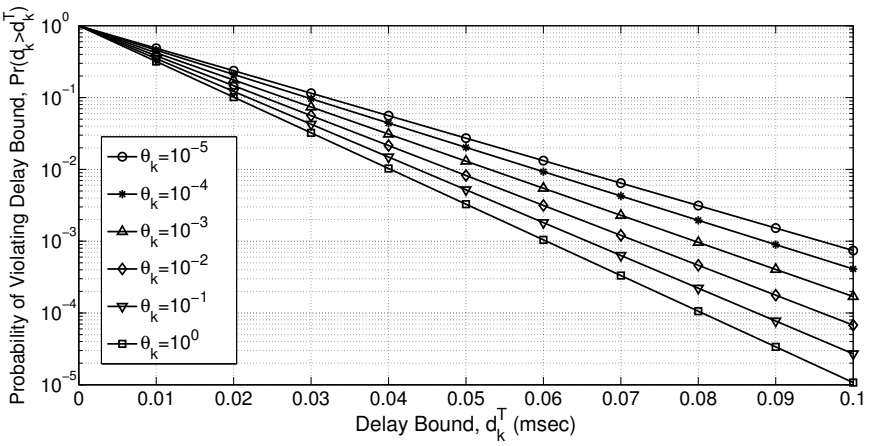


Figure 6.7: Probability of delay violation ($\gamma_{eff,k} = 20$ dB)

Lastly, we discuss the complexity of the resource allocation algortihm based on IWO. It should be noted that IWO is an iterative algorithm and is used for each dual variable associated with the dual function in (6.16). In each iteration for $\boldsymbol{\delta}^*, \boldsymbol{\beta}^*, \boldsymbol{\tau}^*$, and $\boldsymbol{\sigma}^*$, we compute K, KA, L, and K variables respectively. Thus the algorithm has an overall worst case complexity[5] of $\mathcal{O}\left(iter_{max} \cdot \{L + 2K + KA\}\right)$. The performance of IWO algorithm in terms of convergence and computational time has been thoroughly investigated in [12] and [154] through comparison with Genetic Algorithm, Particle Swarm Optimization, Differential Evolution, and other evolutionary algorithms. We assume that the eNB is capable (in terms of processing power) of running the required number of iterations for IWO in one TTI.

6.6 Summary and Concluding Remarks

Energy efficiency requirements have become particularly important in cellular networks with the introduction of MTC. In addition, resource allocation in M2M/H2H co-existence scenarios becomes challenging due to the diverse QoS requirements of M2M and H2H users.

[5]In literature, the complexity of the IWO algorithm is usually given in terms of computational time [12]. To the best of our knowledge, the conventional *Big-O* complexity of IWO does not exist in literature. This is possibly due to the fact the finding the Big-O complexity of IWO is rather cumbersome as each iteration of IWO consists of different operations such as select, sort, and insert operations, which further consist of different arithmetic operations.

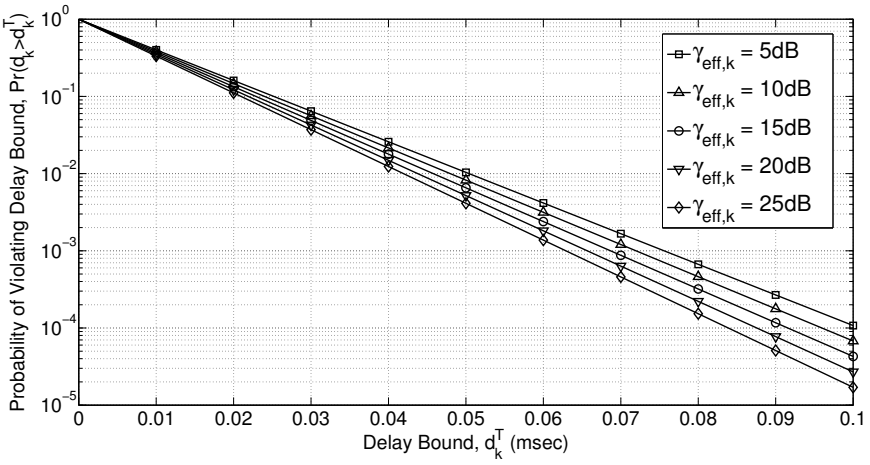


Figure 6.8: Probability of delay violation ($\theta_k = 10^{-1}$)

In this chapter[6], we have investigated energy efficient resource allocation for the uplink of LTE networks under statistical QoS provisioning. The resource allocation problem, which is complicated due to specific power and RB allocation constraints of SC-FDMA, is solved using CDT by first formulating a MIP problem and then converting into a canonical dual problem. We employ the IWO algorithm for obtaining the solution of canonical dual problem. Numerical results have shown that the proposed energy efficient design not only outperforms classical algorithms in terms of energy efficiency but also performs very close to the optimal solution, while satisfying the QoS requirements of different users. Results also show that orthogonal allocation for MTCDs and H2H users is sub-optimal. Moreover, P2P MTCD communication using cellular resources results in intra-cell interference that degrades overall energy efficiency and therefore necessitates the adoption of network-assisted D2D technologies for realizing this type of communication.

[6]Apart from the energy efficient resource allocation problem presented in this chapter, the author has also investigated a power efficient resource allocation problem for the uplink of LTE networks under statistical QoS provisioning [156]. The objective therein is to minimize the sum power in the uplink under QoS requirements of different users and the specific constraints of SC-FDMA. The resource allocation problem is solved using the proposed canonical dual framework and the IWO algorithm. In addition, a heuristic resource allocation algorithm has been developed for a similar resource allocation problem with an average delay constraint [141].

Chapter 7

Error Performance of Diffusion-based Molecular Nano-M2M Networks

7.1 Introduction

Molecular Communication (MC) [15] is a promising communication paradigm for Nano-M2M networks (Nanonetworks), which is inspired by communication of biological systems at cellular level. In MC, molecules are used to encode, transmit, and receive information at the nanoscale. MC can be classified into three different categories based on the type of molecular propagation [68]. In *walkaway-based* MC, the molecules propagate through pre-defined pathways using carrier substances such as molecular motors. In *flow-based* MC, molecules propagate in guided fluidic medium. In *diffusion-based* MC (DMC), molecules propagate through their spontaneous diffusion in a fluidic medium and therefore, the movement is governed by the laws of diffusion. Our focus in this chapter is on DMC as it represents the most general and widespread MC architecture found in nature and also the most widely investigated MC technique to date.

In DMC, attenuation of the molecular concentration as it travels in the environment via the diffusion process is a major problem as it not only limits the range of communication but also yields low rates and reliability. On the other hand, *network coding* has emerged as a promising solution to improve throughput and reliability in traditional wireless networks [157], [158]. Recently, relaying and network coding based solutions have been proposed for molecular nanonetworks. The authors in [159] calculate the capacity of a Calcium signaling based molecular relay channel. Similarly, authors in [160] investigate different relaying schemes from information theoretic perspective. In [161], the authors investigate rate-delay trade-off with network coding in molecular nanonetworks.

In literature, numerous studies have investigated DMC from various aspects. To the best

of our knowledge, error performance of DMC under different kinds of impairments is not well investigated. Due to the simplicity of NMs, complex modulation and detection techniques cannot be used. Recently, a simple *pulse-based* modulation scheme [162] for DMC has been proposed along with two different pulse detection techniques: *energy detection* and *amplitude detection*. One of the main contribution of this chapter is to evaluate the error performance of DMC employing pulse-based modulation. We model the stochastic nature of a simple nanonetwork comprising of a pair of transmitter and receiver NMs and derive closed-form expressions for probability of error using both energy detection and amplitude detection techniques. The proposed error model explicitly accounts for diffusion noise and Inter Symbol Interference (ISI) effects. Through numerical results, we compare the performance of both detection techniques along with investigating the effect of different parameters on error performance. We also evaluate the channel capacity of pulse modulated DMC. To the best of our knowledge, error performance of network coding in DMC has not been investigated before and therefore, another contribution of this chapter. We model the stochastic nature of a simple nanonetwork with a molecular relay (nano-relay) node and derive analytical expressions for probability of error. Numerical results investigate the effect of different parameters on error performance.

7.2 Error Performance of DMC using Pulse-based Modulation

7.2.1 System Model

We consider a simple diffusion-based nanonetwork comprising of a transmitter nanomachine (NM) A and a receiver NM B as shown in Fig. 7.1. The molecular diffusion channel is primarily governed by *Fick's* second law of diffusion [163], which states that the concentration of molecules, denoted by $\mathcal{C}(r,t)$, at location $r = \sqrt{x^2+y^2+z^2}$ and at time t is described by the following equation

$$\frac{1}{D}\frac{\partial \mathcal{C}(r,t)}{\partial t} = \nabla^2 \mathcal{C}(r,t), \tag{7.1}$$

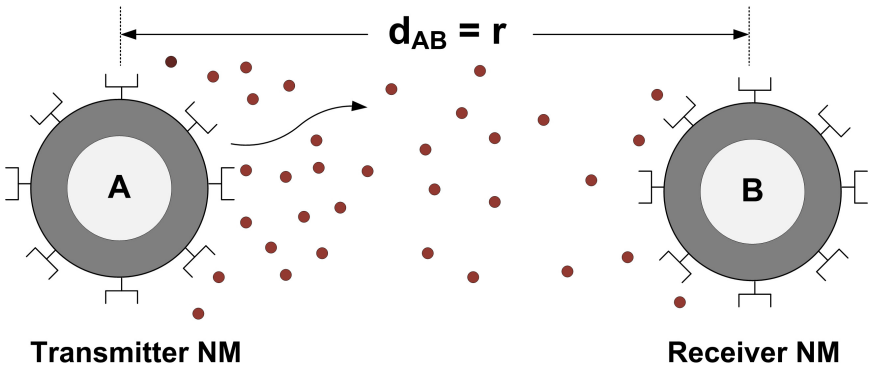


Figure 7.1: A simple diffusion-based molecular nanonetwork

where $D = K_B T_a / 6\pi\eta\ddot{r}$ is the diffusion coefficient of the medium such that K_B is the Boltzmann constant, T_a is the absolute temperature, η is the viscosity constant of the fluid medium, and $\ddot{r}$ is the radius of the molecules. The transmitter NM employs pulse-based modulation scheme, which is based on the solution of (7.1), for information exchange with the receiver NM. According to this scheme, whenever the NM needs to transmit information, it releases a pulse of molecules. This creates a spike in molecular concentration at the transmitter which propagates throughout the medium. If a transmitter releases Q molecules at time instant $t = 0$, the molecular concentration at any point in space and time $t > 0$ is given by

$$\mathcal{C}(r,t) = \frac{Q}{(4\pi Dt)^{3/2}} \cdot e^{-r^2/4Dt}, \tag{7.2}$$

where r is the distance from the transmitter. A plot of (7.2), also known as the *pulse equation* is shown in Fig. 7.2. We observe that the molecular concentration quickly increases until it reaches its maximum value, thereafter gradually decaying over time and forming a long tail due to the effect of diffusion. Note that the maximum molecular concentration is directly proportional to the number of released molecules, Q. Moreover, the maximum concentration is independent of the diffusion coefficient, D. However, a higher value of D reduces the pulse width.

In energy detection, the receiver NM measures the energy of the molecular pulse which is defined as the integral of the molecular concentration over time. Let $\mathcal{E}_p$ denote the pulse

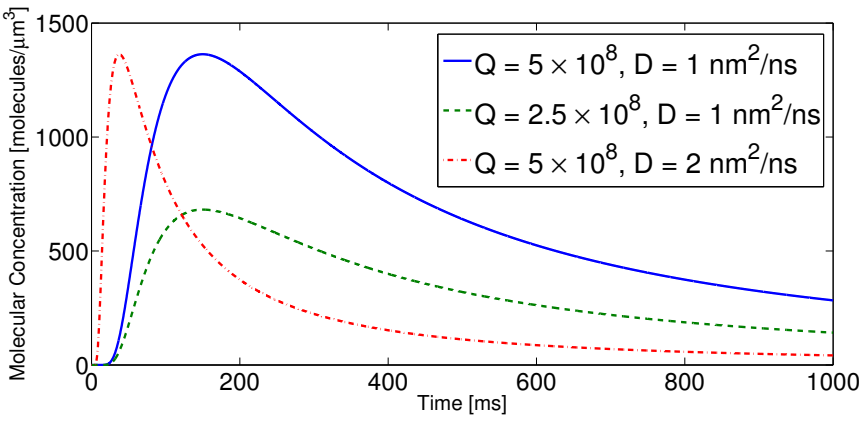


Figure 7.2: Molecular concentration as a function of time, $r = 30\mu$m

energy such that

$$\mathcal{E}_p = \int_0^{T_p} \mathcal{C}(r,t)\mathrm{d}t = \frac{Q}{4\pi D r} erfc\left(\sqrt{\frac{r^2}{4DT_p}}\right), \tag{7.3}$$

where T_p is the *pulse duration* and $erfc(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty e^{-y^2}\mathrm{d}y$ is the *complementary error function*. The received signal is decoded by comparing $\mathcal{E}_p$ with a threshold.

On the other hand, in amplitude detection, the receiver NM measures the variation of local concentration, as given by (7.2), over time. The received signal is decoded by comparing the maximum concentration to a threshold value. The maximum concentration, also known as the pulse amplitude, is given by

$$\mathcal{C}_{max} = \left(\frac{3}{2\pi e}\right)^{3/2} \frac{Q}{r^3} \tag{7.4}$$

7.2.2 Error Performance Analysis of Energy Detection

For simplicity, we consider a binary (2-level) pulse modulation scheme wherein 1 bit per symbol is transmitted. However, the proposed methodology can be extended to higher order modulation schemes as well. Let Q_1 and Q_0 denote the number of released molecules corresponding to bits 1 and 0 respectively (and hence $\mathcal{C}^1(r,t)$ and $\mathcal{C}^0(r,t)$ respectively), with $Q_1 > Q_0$. Let, $E = \mathbf{1}\{\widehat{x}_A \neq x_A\}$ denote the error event, where $\widehat{x}_A$ denotes the decoded symbol at the NM B corresponding to the transmitted message x_A from the NM A.

We define $P_e^{\mathcal{E}}$ as the average probability of error when $\widehat{x}_A$ is in error using energy

detection, which can be calculated as

$$\begin{aligned} P_e^{\mathcal{E}} &= P(E \mid x_A = 0) \cdot P(x_A = 0) + P(E \mid x_A = 1) \cdot P(x_A = 1) \\ &= 0.5 \times \{P(E \mid x_A = 0) + P(E \mid x_A = 1)\}, \end{aligned} \tag{7.5}$$

where it is assumed that both 0 and 1 are equally likely to occur.

Using energy detection, the error probability when $x_A = 0$ is given by

$$P(E \mid x_A = 0) = P\left(\int_0^{T_p} \mathcal{R}(r,t)\mathrm{d}t > \tau\right), \tag{7.6}$$

where τ is the decoding threshold (in terms of pulse energy) and $\mathcal{R}(r,t)$ is the received concentration at the receiver NM. Note that the received concentration is different from the generated concentration due the presence of *diffusion noise*, given by $n(r,t)$. As per [162,164], $n(r,t) \sim \mathcal{N}(0, \sigma^2(r,t))$ i.e., diffusion noise is normal distributed with zero mean and variance of

$$\sigma^2(r,t) = \frac{3}{4\pi\rho^3} \cdot \mathcal{C}(r,t), \tag{7.7}$$

where ρ is the radius of the spherical receiver NM. Therefore, $\mathcal{R}(r,t) = \mathcal{C}^0(r,t) + n(r,t)$. Let, $\mathcal{E}_p^0 = \int_0^{T_p} \mathcal{C}^0(r,t)\mathrm{d}t$ and $\chi_0 = \int_0^{T_p} n(r,t)\mathrm{d}t$. Hence, $P(E \mid x_A = 0)$ is given as follows.

$$P(E \mid x_A = 0) = P\left[\left(\mathcal{E}_p^0 + \chi_0\right) > \tau\right] = P\left[\chi_0 > \tau - \mathcal{E}_p^0\right] \tag{7.8}$$

It should be noted that $n(r,t)$ is actually a random variable (RV). Moreover, the sum of two normal RVs with means μ_a and μ_b and variances of σ_a^2 and σ_b^2 respectively is another normal RV with mean $(\mu_a + \mu_b)$, and variance $(\sigma_a^2 + \sigma_b^2)$ [165]. Therefore, χ_0 is a normal RV with zero mean and variance of $\widetilde{\sigma}_0^2 = \int_0^{T_p} \sigma_0^2(r,t)\mathrm{d}t$, where $\sigma_0^2(r,t) = \frac{3}{4\pi\rho^3} \cdot \mathcal{C}^0(r,t)$.

An important issue in DMC is the Inter Symbol Interference (ISI), which arises due to residue molecules from previously transmitted symbols. Owing to diffusion dynamics, some molecules may arrive after their intended time slot and cause the receiver node to decode the next intended symbol incorrectly. It has been shown in [166] that only the last symbol has significant ISI effect over the current symbol. Note that ISI occurs when a receiver node performs two consecutive decoding operations.

The probability of ISI from a single molecule arriving within a time slot $[T_p, 2T_p]$ can be calculated as $p = \int_{T_p}^{2T_p} f(r,t)\mathrm{d}t$, where $f(r,t)$ denotes the probability density function (PDF) of the travelling time of the molecule from transmitter to receiver and is given by [167]

$$f(r,t) = \frac{r}{\sqrt{4\pi D t^3}} e^{-r^2/4Dt} \tag{7.9}$$

Given a total of Q released molecules, ISI is binomial distributed i.e., $I(r,t) \sim Binomial(Q,p)$. For large Q, the binomial distribution can be approximated with a normal distribution [165]. Hence, $I(r,t) \sim \mathcal{N}\left(Qp, Qp(1-p)\right)$.

With the addition of ISI, (7.8) is modified as follows.

$$P(E \mid x_A = 0) = P\left[\chi_0 + \chi_{ISI} > \tau - \mathcal{E}_p^0\right], \tag{7.10}$$

where χ_{ISI} a normal RV[1] with mean $\breve{\mu}$ and variance $\widetilde{\sigma}_{ISI}^2$ such that

$$\begin{cases} \breve{\mu} = 0.5 \int_{T_p}^{2T_p} (Q_1 + Q_0) p \mathrm{d}t \\ \widetilde{\sigma}_{ISI}^2 = 0.5 \int_{T_p}^{2T_p} (Q_1 + Q_0) p(1-p) \mathrm{d}t \end{cases} \tag{7.11}$$

Let, $\chi_{c_0} = \chi_0 + \chi_{ISI}$ such that $\chi_{c_0} \sim \mathcal{N}\left(\mu_{c_0} = \breve{\mu}, \sigma_{c_0}^2 = \widetilde{\sigma}_0^2 + \widetilde{\sigma}_{ISI}^2\right)$. Therefore, using the PDF of a normal distribution

$$\begin{aligned} P(E \mid x_A = 0) &= \frac{1}{\sqrt{2\pi\sigma_{c_0}^2}} \int_{\tau - \mathcal{E}_p^0}^{\infty} e^{-(\chi_{c_0} - \mu_{c_0})^2 / 2\sigma_{c_0}^2} \mathrm{d}\chi_{c_0} \\ &= \frac{1}{2} erfc\left(\frac{\tau - \mathcal{E}_p^0 - \breve{\mu}}{\sqrt{2}\sigma_{c_0}}\right) \end{aligned} \tag{7.12}$$

Similarly, the error probability when $x_A = 1$ can be calculated as

$$P(E \mid x_A = 1) = P\left(\int_0^{T_p} \mathcal{R}(r,t)\mathrm{d}t \leq \tau\right) = P\left[\chi_1 + \chi_{ISI} \leq \tau - \mathcal{E}_p^1\right], \tag{7.13}$$

where $\mathcal{E}_p^1 = \int_0^{T_p} \mathcal{C}^1(r,t)\mathrm{d}t$, and χ_1 is a normal RV with zero mean and variance of $\widetilde{\sigma}_1^2 = \int_0^{T_p} \sigma_1^2(r,t)\mathrm{d}t$, such that $\sigma_1^2(r,t) = \frac{3}{4\pi\rho^3} \cdot \mathcal{C}^1(r,t)$.

[1] As the symbol causing ISI can be either 0 or 1, the mean and variance are calculated accordingly.

Let $\chi_{c_1} = \chi_1 + \chi_{ISI}$ such that $\chi_{c_1} \sim \mathcal{N}\left(\mu_{c_1} = \breve{\mu}, \sigma_{c_1}^2 = \widetilde{\sigma}_1^2 + \widetilde{\sigma}_{ISI}^2\right)$. Therefore,

$$\begin{aligned} P(E \mid x_A = 1) &= \frac{1}{\sqrt{2\pi\sigma_{c_1}^2}} \int_0^{\tau - \mathcal{E}_p^1} e^{-(\chi_{c_1} - \mu_{c_1})^2 / 2\sigma_{c_1}^2} \mathrm{d}\chi_{c_1} \\ &= \frac{1}{2}\left[erf\left(\frac{\breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right) + erf\left(\frac{\tau - \mathcal{E}_p^1 - \breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right)\right], \end{aligned} \tag{7.14}$$

where $erf(\cdot) = 1 - erfc(\cdot)$ is the *error function*.

Using (7.5), the average probability of error when $\widehat{x}_A$ is in error using energy detection is given by

$$P_e^{\mathcal{E}} = \frac{1}{4}\left[erfc\left(\frac{\tau - \mathcal{E}_p^0 - \breve{\mu}}{\sqrt{2}\sigma_{c_0}}\right) + erf\left(\frac{\breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right) + erf\left(\frac{\tau - \mathcal{E}_p^1 - \breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right)\right] \tag{7.15}$$

7.2.3 Error Performance Analysis of Amplitude Detection

We define $P_e^{\mathcal{A}}$ as the average probability of error when $\widehat{x}_A$ is in error using amplitude detection. Note that $P_e^{\mathcal{A}}$ can be calculated in a similar way as described for $P_e^{\mathcal{E}}$.

Using amplitude detection, the error probability when $x_A = 0$ is given by

$$P(E \mid x_A = 0) = P\left(\mathcal{R}(r,t) > \tau^a\right), \tag{7.16}$$

where τ^a is the decoding threshold (in terms of pulse amplitude). Incorporating the effect of diffusion noise and ISI,

$$P(E \mid x_A = 0) = P\left[\lambda_0 + \lambda_{ISI} > \tau^a - \mathcal{A}_p^0\right], \tag{7.17}$$

where $\mathcal{A}_p^0 = \mathcal{C}_{max}^0$, $\lambda_0 \sim \mathcal{N}\left(0, \ddot{\sigma}_0^2(r,t) = \frac{3}{4\pi\rho^3} \cdot \mathcal{A}_p^0\right)$, and $\lambda_{ISI} \sim \mathcal{N}(\ddot{\mu}, \ddot{\sigma}_{ISI}^2)$ such that

$$\begin{cases} \ddot{\mu} = 0.5(Q_1 + Q_0)p \\ \ddot{\sigma}_{ISI}^2 = 0.5(Q_1 + Q_0)p(1-p) \end{cases} \tag{7.18}$$

Let, $\lambda_{c_0} = \lambda_0 + \lambda_{ISI}$ such that $\lambda_{c_0} \sim \mathcal{N}\left(\bar{\mu}_{c_0} = \ddot{\mu}, \bar{\sigma}^2_{c_0} = \ddot{\sigma}^2_0(r,t) + \ddot{\sigma}^2_{ISI}\right)$. Therefore,

$$\begin{aligned} P(E \mid x_A = 0) &= \frac{1}{\sqrt{2\pi\bar{\sigma}^2_{c_0}}} \int_{\tau^a - \mathcal{A}^0_p}^{\infty} e^{-(\lambda_{c_0} - \bar{\mu}_{c_0})^2 / 2\bar{\sigma}^2_{c_0}} \mathrm{d}\lambda_{c_0} \\ &= \frac{1}{2} erfc\left(\frac{\tau^a - \mathcal{A}^0_p - \ddot{\mu}}{\sqrt{2}\bar{\sigma}_{c_0}}\right) \end{aligned} \tag{7.19}$$

Similarly, the error probability when $x_A = 1$ can be calculated as

$$P(E \mid x_A = 1) = P\left(\mathcal{R}(r,t) \leq \tau^a\right) = P\left[\lambda_1 + \lambda_{ISI} \leq \tau^a - \mathcal{A}^1_p\right], \tag{7.20}$$

where $\mathcal{A}^1_p = \mathcal{C}^1_{max}$ and $\lambda_1 \sim \mathcal{N}\left(0, \ddot{\sigma}^2_1(r,t) = \frac{3}{4\pi\rho^3} \cdot \mathcal{A}^1_p\right)$.

Let, $\lambda_{c_1} = \lambda_1 + \lambda_{ISI}$ such that $\lambda_{c_1} \sim \mathcal{N}\left(\bar{\mu}_{c_1} = \ddot{\mu}, \bar{\sigma}^2_{c_1} = \ddot{\sigma}^2_1(r,t) + \ddot{\sigma}^2_{ISI}\right)$. Therefore,

$$\begin{aligned} P(E \mid x_A = 1) &= \frac{1}{\sqrt{2\pi\bar{\sigma}^2_{c_1}}} \int_{0}^{\tau^a - \mathcal{A}^1_p} e^{-(\lambda_{c_1} - \bar{\mu}_{c_1})^2 / 2\bar{\sigma}^2_{c_1}} \mathrm{d}\lambda_{c_1} \\ &= \frac{1}{2}\left[erf\left(\frac{\ddot{\mu}}{\sqrt{2}\bar{\sigma}_{c_1}}\right) + erf\left(\frac{\tau^a - \mathcal{A}^1_p - \ddot{\mu}}{\sqrt{2}\bar{\sigma}_{c_1}}\right)\right] \end{aligned} \tag{7.21}$$

Hence, the average probability of error when $\widehat{x}_A$ is in error using amplitude detection is given by

$$P^{\mathcal{A}}_e = \frac{1}{4}\left[erfc\left(\frac{\tau^a - \mathcal{A}^0_p - \ddot{\mu}}{\sqrt{2}\bar{\sigma}_{c_0}}\right) + erf\left(\frac{\ddot{\mu}}{\sqrt{2}\bar{\sigma}_{c_1}}\right) + erf\left(\frac{\tau^a - \mathcal{A}^1_p - \ddot{\mu}}{\sqrt{2}\bar{\sigma}_{c_1}}\right)\right] \tag{7.22}$$

7.3 Channel Capacity of Pulse-Modulated DMC

We consider a binary channel with inputs $[X_0, X_1]$ and the corresponding outputs $[Y_0, Y_1]$. Using the error probability expressions for energy detection and amplitude detection, and selecting ideal threshold values, the channel capacity (C) can be calculated as the maximum of the mutual information, and is given as follows [168].

$$C = \max_{\tau/\tau^a} I(X,Y), \tag{7.23}$$

where $I(X,Y)$ is given by (7.25) such that

$$\begin{cases} P(X_0) = P(X_1) = 0.5 = P(Y_0) = P(Y_1) \\ P(Y_1|X_0) = P(E \mid x_A = 0) \\ P(Y_0|X_1) = P(E \mid x_A = 1) \\ P(Y_0|X_0) = 1 - P(E \mid x_A = 0) \\ P(Y_1|X_1) = 1 - P(E \mid x_A = 1) \\ P(X_i, Y_i) = P(X_i)P(Y_i|X_i) \quad i \in 0, 1 \end{cases} \tag{7.24}$$

$$\begin{aligned} I(X,Y) = P(X_0,Y_0)\log_2\frac{P(Y_0|X_0)}{P(Y_0)} + P(X_0,Y_1)\log_2\frac{P(Y_1|X_0)}{P(Y_1)} \\ + P(X_1,Y_0)\log_2\frac{P(Y_0|X_1)}{P(Y_0)} + P(X_1,Y_1)\log_2\frac{P(Y_1|X_1)}{P(Y_1)} \end{aligned} \tag{7.25}$$

7.4 Error Performance of Networking Coding in Diffusion-based Molecular Nanonetworks

7.4.1 System Model

We consider a simple diffusion-based nanonetwork comprising of NMs (nodes) A and B and a nano-relay node as shown in Fig. 7.3. We assume that nodes employ pulse-based modulation scheme with energy detection. Further, the NMs A and B are not in communication range of each other but in the communication range of the nano-relay. Without network coding, NM A sends its message x_A to the nano-relay which forwards it to the NM B. Similarly, NM B sends its message x_B to the nano-relay which forwards it to the NM A. Thus, we have four transmissions (in four timeslots) in total. When network coding is employed, the same information can be exchanged in three transmissions. In the first time slot, NM A sends x_A to the nano-relay. Similarly, NM B transmits x_B to the nano-relay in the second time slot. The nano-relay XORs[2] the received messages ($x_R = x_A \oplus x_B$) and sends the combined message to NMs A and B in the third time slot. Since NM A has *a priori* information of x_A, it can decode x_B by performing the operation $x_B = x_R \oplus x_A$. Similarly,

[2]We assume that the nano-relay is equipped with an XOR gate implemented at molecular level by pseudorotaxane [169].

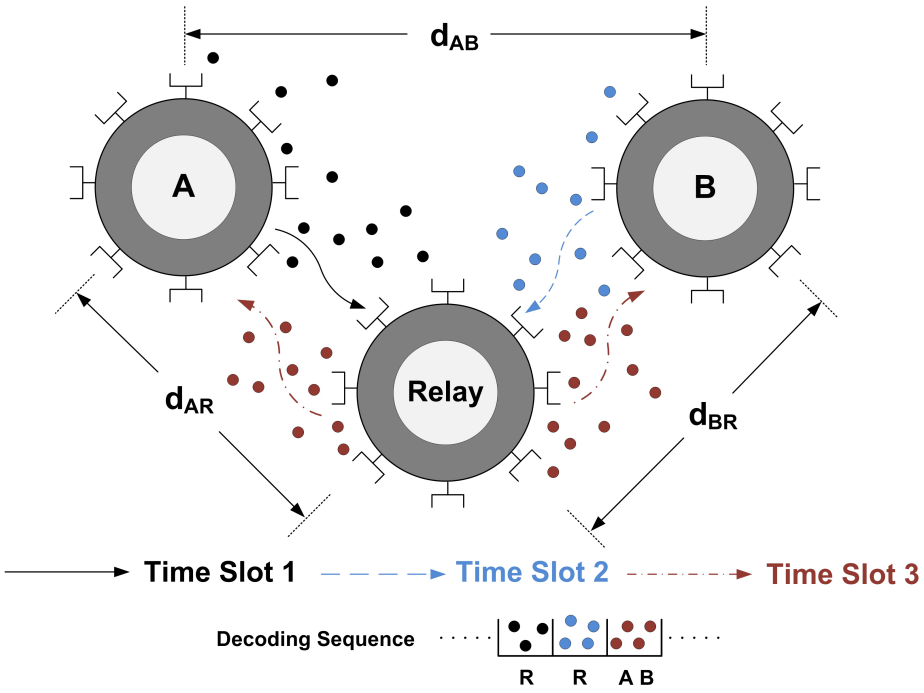


Figure 7.3: A simple nanonetwork with network coding operation

the NM B decodes x_A through $x_A = x_R \oplus x_B$.

7.4.2 Error Performance Analysis

We consider a binary (2-level) pulse modulation scheme. However, the analysis can be extended to higher order modulation schemes as well. Next, we describe the possible error scenarios in the network coded system under consideration. Let, $\widehat{x}_A$ and $\widehat{x}_B$ denote the decoded symbols at the nano-relay corresponding to x_A and x_B respectively. Similarly, let $\widehat{x}_R$ denote the decoded symbol at nodes A and B, corresponding to $x_R = x_A \oplus x_B$. Let, $E = \mathbf{1}\{\widehat{x}_m \neq x_m, m \in (A, B, R)\}$ denote the error event. An error occurs in two different cases; (i) when x_R is in error, and (ii) when $\widehat{x}_R$ is in error. In the first scenario, x_R is in error if either $\widehat{x}_A$ or $\widehat{x}_B$ is in error. Note that due to the property of XOR function, no error occurs if both $\widehat{x}_A$ and $\widehat{x}_B$ are in error.

Recall that ISI occurs when a receiver node performs two consecutive decoding operations. As seen from Fig. 7.3, ISI occurs when two consecutive decoding operations are performed by the nano-relay. Hence, $\widehat{x}_B$ will be affected by ISI due to transmitted

symbol from NM A in the previous time slot.

We define $\bar{P}_e^A$ as the average probability of error when $\widehat{x}_A$ is in error, which is given by

$$\begin{aligned}\bar{P}_e^A &= P(E \mid x_A = 0) \cdot P(x_A = 0) + P(E \mid x_A = 1) \cdot P(x_A = 1) \\ &= 0.5 \times \{P(E \mid x_A = 0) + P(E \mid x_A = 1)\}\end{aligned} \tag{7.26}$$

where it is assumed that both 0 and 1 occur with equal probability.

Since the NMs and the nano-relay employ energy detection, the error performance can be calculated in a similar way as described in Section 7.2.2. Since, $\widehat{x}_A$ is not affected by ISI, the error probability when $x_A = 0$ is given by

$$P(E \mid x_A = 0) = P\left[\chi_0 > \tau - \mathcal{E}_p^0\right] = \frac{1}{2} erfc\left(\frac{\tau - \mathcal{E}_p^0}{\sqrt{2}\widetilde{\sigma}_0}\right) \tag{7.27}$$

Similarly, the error probability when $x_A = 1$ is given by

$$P(E \mid x_A = 1) = P\left[\chi_1 \leq \tau - \mathcal{E}_p^1\right] = \frac{1}{2} erf\left(\frac{\tau - \mathcal{E}_p^1}{\sqrt{2}\widetilde{\sigma}_1}\right) \tag{7.28}$$

Note that $\mathcal{E}_p^0$, $\mathcal{E}_p^1$, τ, χ_0, χ_1, $\widetilde{\sigma}_0$ and $\widetilde{\sigma}_1$ are the same as in Section 7.2.2. Using (7.26), the overall probability of error when $\widehat{x}_A$ is in error is given by

$$\bar{P}_e^A = \frac{1}{4}\left[erfc\left(\frac{\tau - \mathcal{E}_p^0}{\sqrt{2}\widetilde{\sigma}_0}\right) + erf\left(\frac{\tau - \mathcal{E}_p^1}{\sqrt{2}\widetilde{\sigma}_1}\right)\right] \tag{7.29}$$

Let $\bar{P}_e^B$ denote the average error probability when $\widehat{x}_B$ is in error. Since the NM B employs the same modulation scheme as the NM A, P_e^B can be calculated in a similar way as described for P_e^A, with additionally accounting for the ISI.

The expressions for error probability when $x_B = 0$ and $x_B = 1$ are given by (7.30) and (7.31) respectively.

$$P(E \mid x_B = 0) = P\left[\chi_0 + \chi_{ISI} > \tau - \mathcal{E}_p^0\right] = \frac{1}{2} erfc\left(\frac{\tau - \mathcal{E}_p^0 - \breve{\mu}}{\sqrt{2}\sigma_{c_0}}\right) \tag{7.30}$$

$$P(E \mid x_B = 1) = P\left[\chi_1 + \chi_{ISI} \leq \tau - \mathcal{E}_p^1\right] = \frac{1}{2}\left[erf\left(\frac{\breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right) + erf\left(\frac{\tau - \mathcal{E}_p^1 - \breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right)\right] \tag{7.31}$$

Hence, the overall probability of error when $\widehat{x}_B$ is in error is given by

$$\bar{P}_e^B = \frac{1}{4}\left[erfc\left(\frac{\tau - \mathcal{E}_p^0 - \breve{\mu}}{\sqrt{2}\sigma_{c_0}}\right) + erf\left(\frac{\breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right) + erf\left(\frac{\tau - \mathcal{E}_p^1 - \breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right)\right]. \tag{7.32}$$

Therefore, the overall error probability when x_R is in error is given by $P_e^R = \bar{P}_e^A + \bar{P}_e^B$. Next, we consider the second case of error wherein $\widehat{x}_R$ is in error. In this case, the error occurs at either NM A or NM B while decoding x_R. Let $P_e^{\widehat{R}}$ denote the probability when $\widehat{x}_R$ is in error. Since $\widehat{x}_R$ can be either 0 or 1, $P_e^{\widehat{R}}$ is given by $P_e^{\widehat{R}} = 0.5 \times (P(E \mid \widehat{x}_R = 0) + P(E \mid \widehat{x}_R = 1))$. Note that $\widehat{x}_R$ is not affected by ISI. Therefore, $P_e^{\widehat{R}} = \bar{P}_e^A$.

Hence, the overall error probability in a network coded molecular nanonetwork can be calculated as $P_e^{NC} = P_e^R + 2P_e^{\widehat{R}} = 3\bar{P}_e^A + \bar{P}_e^B$, which is given by (7.33).

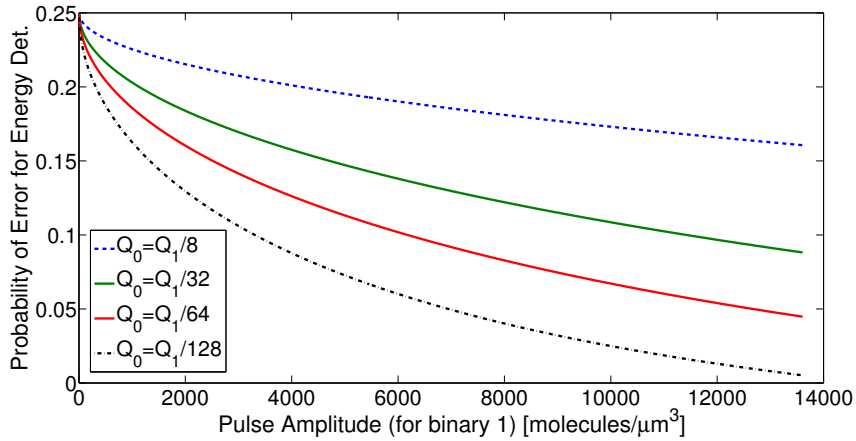
$$\begin{aligned} P_e^{NC} = {} & \frac{3}{4}\left[erfc\left(\frac{\tau - \mathcal{E}_p^0}{\sqrt{2}\widetilde{\sigma}_0}\right) + erf\left(\frac{\tau - \mathcal{E}_p^1}{\sqrt{2}\widetilde{\sigma}_1}\right)\right] \\ & + \frac{1}{4}\left[erfc\left(\frac{\tau - \mathcal{E}_p^0 - \breve{\mu}}{\sqrt{2}\sigma_{c_0}}\right) + erf\left(\frac{\breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right) + erf\left(\frac{\tau - \mathcal{E}_p^1 - \breve{\mu}}{\sqrt{2}\sigma_{c_1}}\right)\right] \end{aligned} \tag{7.33}$$

7.5 Numerical Results

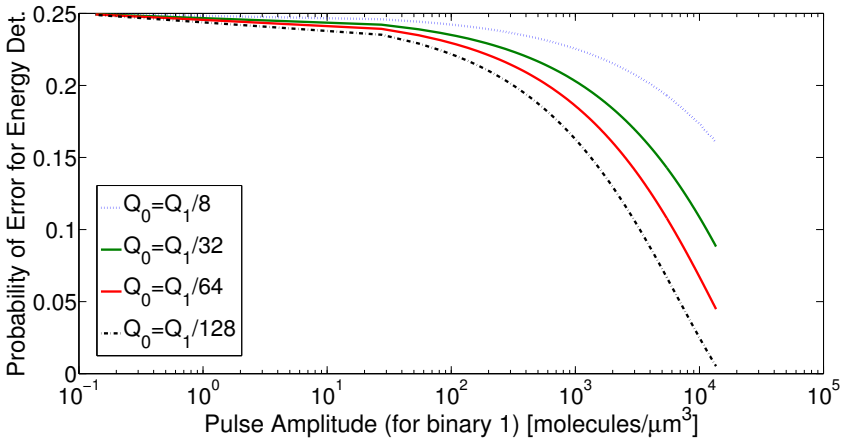
7.5.1 Error Performance of Energy and Amplitude Detection Techniques

For a fair comparison of error performance, we assume that the time slot duration for decoding operation is the same for both energy and amplitude detection techniques and is equal to the pulse duration, T_p. In other words, we assume a fixed transmission rate for both detection techniques. We use the pulse amplitude, given by (7.4), for comparing the error performance of both detection techniques.

Fig. 7.4a shows the error probability for energy detection against the pulse amplitude



(a)



(b)

Figure 7.4: (a) Probability of error for energy detection against the pulse amplitude ($D = 3\text{nm}^2/\text{ns}, T_p = 10^4\text{s}, \rho = 0.4\mu\text{m}, \tau = 0.5 \times [\mathcal{E}_p^0 + \mathcal{E}_p^1], r = 3\mu\text{m}$), (b) Semi-logarithmic plot of error probability for energy detection against the pulse amplitude (similar parameters as in (a)).

for a binary 1. We observe that the error probability is dependent on the pulse amplitude and the molecular level difference of Q_0 and Q_1. The error probability decreases exponentially as pulse amplitude increases due to increase in decision domain for decoding. With a similar reasoning, increasing the molecular level difference improves the error performance. From Fig. 7.4b, we also note that the error performance follows a similar trend as followed by the error probability of binary *Pulse Amplitude Modulation* against *Signal-to-Noise Ratio* in conventional wireless communication.

As shown in Fig. 7.5, the error probability for energy detection increases as the diffusion

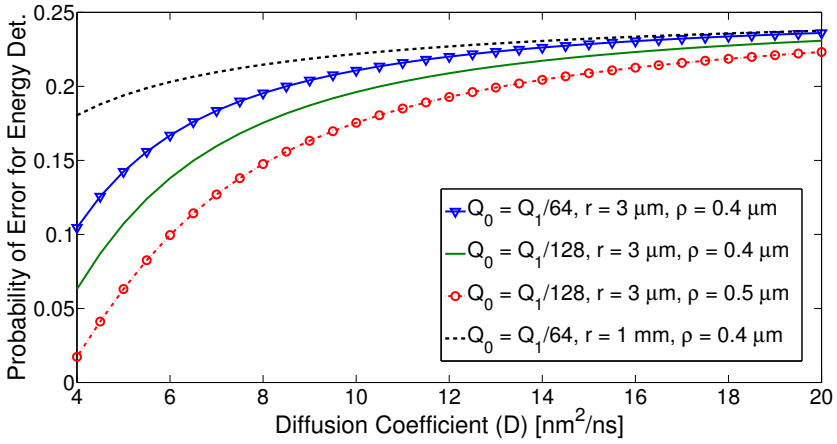


Figure 7.5: Probability of error for energy detection against the diffusion coefficient ($Q_1 = 5 \times 10^6, T_p = 10^4$s, $\tau = 0.5 \times [\mathcal{E}_p^0 + \mathcal{E}_p^1]$)

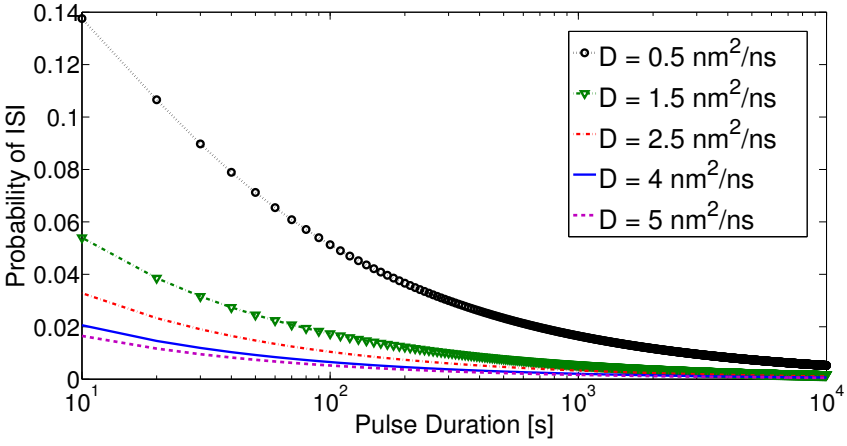
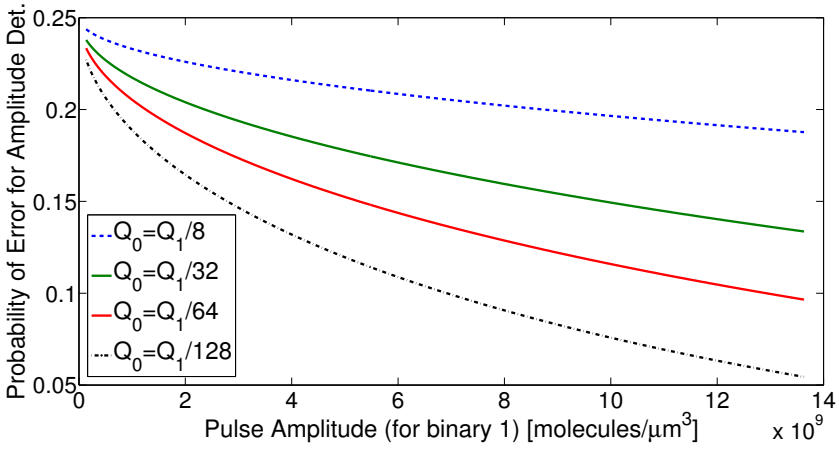
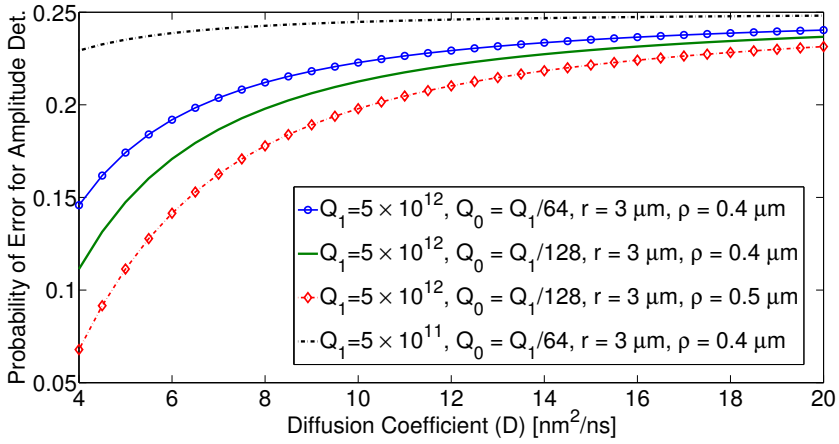


Figure 7.6: Probability of ISI against the pulse duration ($r = 10\mu$m)

coefficient (D) increases. This is because molecules diffuse more freely into the medium as a result of which the molecular pulse decays quickly. The error probability reduces as the radius of receiver NM (ρ) increases due to reduced variance of diffusion noise (reduced noise power). Moreover, the error probability increases as the distance between NMs and the nano-relay increases.



(a)



(b)

Figure 7.7: (a) Probability of error for amplitude detection against the pulse amplitude ($D = 3\text{nm}^2/\text{ns}, T_p = 10^4\text{s}, \rho = 0.4\mu\text{m}, \tau^a = 0.5 \times [\mathcal{A}_p^0 + \mathcal{A}_p^1], r = 3\mu\text{m}$), (b) Probability of error for amplitude detection against the diffusion coefficient ($T_p = 10^4\text{s}, \tau^a = 0.5 \times [\mathcal{A}_p^0 + \mathcal{A}_p^1]$)

Fig. 7.6 shows the probability of ISI[3] from a single molecule, given by p, against the pulse duration (T_p). We note that p decreases as T_p increases. At sufficiently large T_p, the

[3]Our analysis shows that $p = K_1\sqrt{\frac{\pi}{K_2}}\left[erfc\left(\sqrt{\frac{K_2}{2T_p}}\right) - erfc\left(\sqrt{\frac{K_2}{T_p}}\right)\right]$, where $K_1 = r/\sqrt{4\pi D}$ and $K_2 = r^2/4D$.

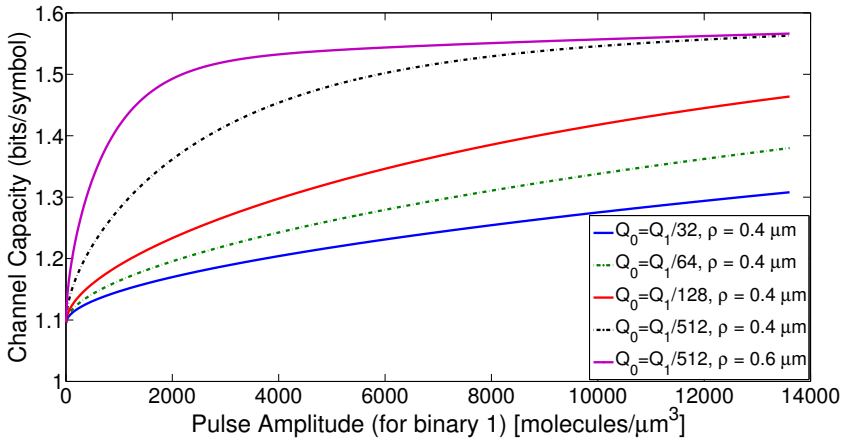
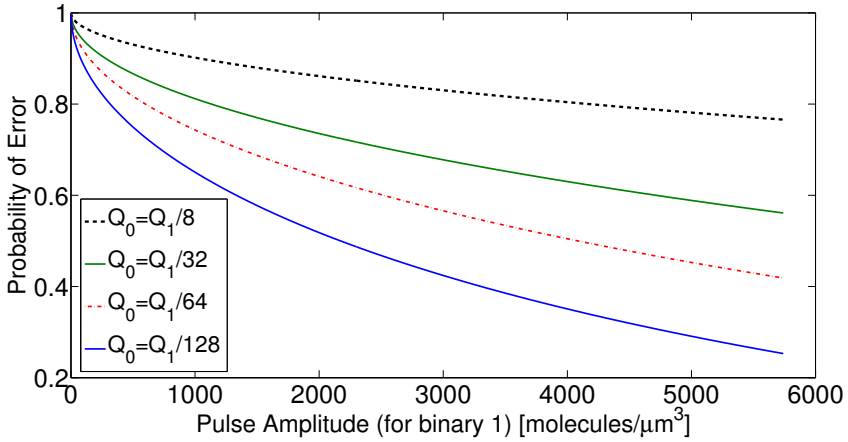


Figure 7.8: Channel capacity for energy detection against the pulse amplitude ($D = 3\text{nm}^2/\text{ns}, T_p = 10^4\text{s}, r = 3\mu\text{m}, \tau = 0.5 \times [\mathcal{E}_p^0 + \mathcal{E}_p^1]$)

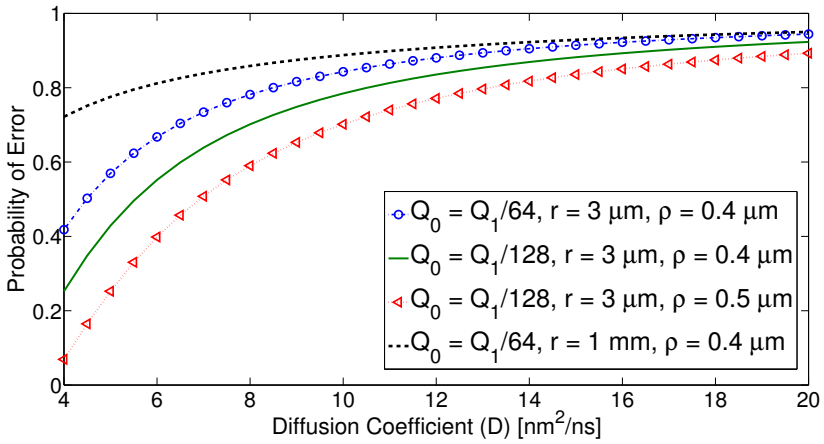
effect of ISI is negligible. This is due to the fact that T_p dictates the transmission rate in DMC. Smaller values of T_p correspond to higher transmission rate that makes the molecular pulses more susceptible to ISI. From another perspective using (7.2), a shorter T_p results in a longer tail of molecules from previous time slot arriving in the current time slot, which results in higher ISI. It should be noted that for a fixed distance and T_p, increasing the diffusion coefficient (D) reduces p as the molecule propagates more freely in the medium, reducing the probability of reaching a receiver node. Note that both detection techniques are equally affected by ISI under the assumption of fixed transmission rate for both schemes.

The results in Fig. 7.7a and 7.7b show the error probability for amplitude detection against the pulse amplitude (for binary 1) and diffusion coefficient (D). The error performance follows a similar trend as followed by energy detection and shows dependence on similar parameters. However, note that for fixed distance, amplitude detection requires much higher concentration of molecules to achieve similar error performance as energy detection. Quantifying this, we note that amplitude detection requires up to 10^6 times more released molecules to achieve similar error performance as energy detection. Our error performance model confirms that energy detection is more suitable for transmission over longer distances as predicted in [162]. The worse performance of amplitude detection is due to the fact that pulse amplitude in DMC scales $\mathcal{O}(1/r^3)$ compared to pulse energy which scales $\mathcal{O}(1/r)$.

Fig. 7.8 evaluates the channel capacity (in bits per symbol) using energy detection. As



(a)



(b)

Figure 7.9: Probability of error for network coding against (a) pulse amplitude ($D = 3\text{nm}^2/\text{ns}, T_p = 10^4\text{s}, \rho = 0.4\mu\text{m}, \tau = 0.5 \times [\mathcal{E}_p^0 + \mathcal{E}_p^1], d_{AR} = d_{BR} = r = 3\mu\text{m}$), (b) diffusion coefficient ($Q_1 = 5 \times 10^6, T_p = 10^4\text{s}, \tau = 0.5 \times [\mathcal{E}_p^0 + \mathcal{E}_p^1]$)

for error probability, the channel capacity depends on the pulse amplitude and the molecular level difference of Q_1 and Q_0. Increasing the molecular level difference significantly improves the achievable capacity. Moreover, for similar level of released molecules, reducing the variance of diffusion noise (through increasing the radius (ρ) of NM) improves the channel capacity, particularly at low pulse amplitude levels. Since amplitude detection requires much higher concentration of released molecules to achieve similar performance as energy detection over a fixed distance, the channel capacity using amplitude detection for similar pulse amplitude values will be much lower.

7.5.2 Error Performance of Network Coding

We assume that the distance of nano-relay from NMs A and B is equal. The results in Fig. 7.9 show the overall error performance of network coding against the pulse amplitude and the diffusion coefficient (D). The error probability decreases exponentially as pulse amplitude increases due to increase in decision domain for decoding. With a similar reasoning, increasing the molecular level difference of Q_0 and Q_1 improves the error performance.

As shown in Fig. 7.9b, the error probability increases as the diffusion coefficient (D) increases. This is due to the fact that the molecular pulse decays quickly as the molecules are able to diffuse more freely into the medium. The error probability reduces as the radius of receiver NM (ρ) increases due to reduced variance of diffusion noise. Moreover, the error probability increases as the distance between NMs and the nano-relay increases.

7.6 Summary and Concluding Remarks

DMC is a promising technique for nano-M2M networks. In this chapter, we have evaluated the error performance of DMC using pulse-based modulation scheme under two different pulse detection techniques; energy detection and amplitude detection. Closed-form expressions for error probability have been derived in the presence of diffusion noise and ISI. The channel capacity of pulse modulated DMC has been evaluated as well. Using the error performance model, error probability expressions for a network coded molecular nano-M2M network have been derived as well. Numerical results show that the error probability depends on pulse amplitude, probability of ISI, diffusion coefficient of the medium, and the variance of diffusion noise. Performance evaluation also demonstrates that energy detection is better suited for transmission over longer distances. The error performance model can be used to obtain insights on reliable communication range of DMC depending on the design of NMs.

Chapter 8

Conclusions and Future Work

8.1 Concluding Remarks

M2M communications acts as an enabling technology for the practical realization of not just the Internet-of-Things (IoT), but also the Internet-of-Nano-Things and the Internet-of-Energy, and ultimately leads to the creation of what is known as the Internet-of-Everything. M2M communications will revolutionize every aspect of present day life by creating smart homes, smart grids, smart transportation, smart buildings, and smart cities.

It was the aim of this thesis to investigate different challenges related to M2M communications. The thesis not only made original contributions to the research community but also opened up interesting areas for future research. Since each chapter of the thesis addressed an independent research problem, the main contributions of the thesis are summarized below along with concluding remarks to present an overall picture of the research conducted.

M2M communications will be realized through a range of technologies and networks. Therefore, the contributions of this thesis cover different aspects of M2M communications. For the first time, the use of cognitive radio technology in capillary M2M networks has been investigated from a protocol stack perspective. Firstly, a centralized cognitive MAC protocol has been proposed that uses a specialized frame structure for co-existence with the primary network along with meeting the utility requirements of the secondary M2M network. The protocol exhibits good scalability characteristics. Throughput of up to 60% or more can be achieved under moderate device density of $80 \sim 100$ devices, depending upon the level of sensed primary signal and the detection probability threshold, while keeping interference to

incumbents to a minimum and satisfying the energy efficiency requirements of the secondary network to an appreciable extent.

Secondly, a distributed cognitive MAC protocol, termed as CRB-MAC, has been proposed under consideration of high energy efficiency and reliability requirements of M2M devices operating in challenging wireless environments. CRB-MAC employs preamble sampling and receiver-based forwarding techniques for providing high energy efficiency and reliability. The protocol is also enhanced to cater for delay sensitive traffic as well as to mitigate the performance degradation under periodic spectrum sensing state of different nodes. Performance evaluation demonstrates that CRB-MAC enhances energy and delay performance by reducing the number of retransmissions. Additionally, reliability of up to 50% can be achieved depending on the number of receivers (deployment density). Hence CRB-MAC provides a viable solution for lossy environments.

Thirdly, an enhanced RPL based routing protocol, termed as CORPL, has been proposed for cognitive radio equipped smart grid AMI networks. In order to ensure spatial and temporal protection to primary users along with meeting the utility requirements of the secondary network, CORPL utilizes an opportunistic forwarding approach with a simple overhearing based coordination scheme. Performance evaluation shows that CORPL improves the reliability of the secondary network by up to 50%, while reducing the probability of interference to the primary network as well as the probability of deadline violation for delay sensitive traffic to a significant extent.

For the first time in cellular M2M, a resource allocation algorithm for the uplink of LTE networks in M2M/H2H co-existence scenarios has been proposed with special emphasis on energy efficiency and statistical QoS requirements. The resource allocation problem, which is particularly complicated due to the specific constraints of SC-FDMA, has been solved using canonical duality theory and invasive weed optimization algorithm. Numerical results demonstrate that the proposed energy efficient design not only outperforms classical algorithms in terms of energy efficiency but also performs very close to the optimal solution, while satisfying the QoS requirements of different users. It is also concluded that orthogonal allocation for MTCDs and H2H users is sub-optimal. In addition, P2P MTCD communication using cellular resources results in intra-cell interference that degrades

overall energy efficiency and therefore, necessitates the adoption of network-assisted D2D technologies for realizing this type of communication.

Nano-M2M is the new frontier in M2M communications. For the first time, a comprehensive error performance model for pulse modulated diffusion-based molecular nano-M2M networks has been developed under diffusion noise and ISI impairments. Error performance of network coding in nano-M2M networks has been evaluated as well. The error probability depends on the pulse amplitude, diffusion coefficient of the medium, variance of the diffusion noise, and the probability of ISI based on slot duration for the decoding operation. The model provides insights on the reliable range of molecular communication, which is useful in the design of receiver nanomachines.

Based on the overall picture of the research conducted in this thesis, additional conclusions can be drawn, some of which are presented as follows.

- Cognitive radio technology will play an important role in realizing the vision of Internet-of-Things. Therefore, cognitive radio aware protocols are required at different layers of the protocol stack that not only fulfil the requirements of M2M devices but also provide a low cost dynamic spectrum access solution.
- A single M2M design cannot possibly fulfil the requirements of all M2M applications i.e., protocols for M2M cannot be generalized from an application perspective. Hence, protocol design for M2M networks must be application driven. For example, smart grid applications have been the main focus throughout this thesis.
- M2M networks require new solutions and protocols. However, for easy integration with existing networking infrastructure, conventional protocols require a revisit mainly from energy efficiency, reliability, and scalability perspectives. One example is the CORPL protocol that provides IP connectivity for cognitive radio equipped M2M networks. Another example is the low-power Wi-Fi [25], which has been adapted for M2M applications through the addition of a power saving mode to IEEE 802.11 standard.
- QoS requirements in M2M are mostly application specific. The resource constrained nature of M2M devices makes QoS provisioning a challenging task for varying data delivery models such as periodic, event-driven, query-initiated, etc. Therefore, context

aware protocols are required for M2M, in order to meet the diverse service requirements as well as to provide differentiated QoS.

- Nano-M2M communication will play an important role in realizing the vision of Internet-of-Nano-Things. Molecular communication will be widely used especially in biomedical applications. However, protocols for nano-M2M networks require simplicity rather than sophistication.

It can be easily inferred that the research on M2M communications is far from complete. This thesis covers some of the key issues in M2M communications and provides the foundations for future research. A number of challenges remain that need to be addressed. The following sections highlight some research directions for future researchers.

8.2 Future Work for Capillary M2M

8.2.1 PRMA-based Cognitive M2M Communications with Multiple Device Domains

It would be interesting to investigate the PRMA-based cognitive MAC protocol, presented in chapter 3, with multiple device domains (multiple M2M networks). In this case, each device domain has a centralized controller, which is the MTCG. The primary challenge is the co-existence of multiple secondary M2M networks as each network may have different service requirements. A simple MAC layer co-existence solution is the frame scheduling for each M2M network based on its priority. However, this requires a common control channel for communication among MTCGs belonging to different M2M networks. Apart from this, certain aspects of protocol operation require further investigation such as the design of reservation cycle. Moreover, cooperative spectrum sensing techniques [170] can be employed to improve the channel detection performance by exploiting the correlation in the sensed information of multiple MTCGs.

8.2.2 Protocol Design for Smart Grid Networks under Power Systems Dynamics

Since smart grid consists of both power and communication layers [32], it is important to consider the characteristics of power systems in protocols designed for the smart grid. In chapter 3, a scheduling algorithm for smart meters has been proposed that takes into account the power dynamics of each smart meter. It would be interesting to investigate other protocols from the perspective of power systems. For example, in CORPL protocol (presented in chapter 5), two different routing classes have been proposed, mainly from communication perspective. Therefore, a new class of CORPL protocol is required, that is specifically designed under the dynamics of power systems. This requires the introduction of power market mechanisms like locational marginal price [110] as well as models for power load variation at each node. The opportunistic forwarding procedure needs to be modified accordingly. Similarly, CRB-MAC protocol (presented in chapter 4) can be investigated from power systems perspective. Specifically, in CRB-MAC, the receiver-based forwarding process needs to be modified such that the node that provides the best performance under power dynamics is the winner.

8.2.3 Cross-Layer Protocol Design

The main focus in this thesis has been on single-layer protocol design. Joint design of different layers (cross-layer design) can improve the performance of M2M networks through joint optimization of parameters at different layers. Therefore, it would be interesting to investigate joint design of RPL with PRMA-based or TDMA-based scheduling in multi-hop M2M networks for both cognitive and non-cognitive scenarios. Similarly, joint design of RPL/CORPL and CRB-MAC protocol (presented in chapter 4) can be explored. It is particularly important to investigate the performance of proposed protocols under the dynamics of transport layer. This may provide useful insights on designing low overhead transport protocols for M2M networks.

8.2.4 New Primary User Activity Models

For analytical modeling and performance analysis in chapters 3, 4, and 5, the primary user (PU) activity (channel usage) is modeled by a two state independent and identically distributed (i.i.d) random process with exponentially distributed busy (on) and idle (off) periods. This assumption of exponentially distributed on-off periods, which has been widely adopted in the context of cognitive radio networks, was originally used for characterizing human voice communications. However, current wireless and cellular networks carry a lot of bursty data traffic. Moreover, future cognitive radio networks will also use the TV bands where the incumbents may or may not be TV transmitters. Therefore, the conventional PU activity model of exponentially distributed on-off periods requires a revisit.

8.3 Future Work for Cellular M2M

8.3.1 Packet Scheduling Algorithms

3GPP LTE networks are becoming increasingly popular for M2M communications. Existing LTE scheduling schemes are designed and optimized to cater for a limited number of users demanding high data rate services. On the contrary, M2M traffic load is expected to be generated in terms of small data transmissions from a large number of M2M devices, which are characterized by diverse QoS requirements (in terms of delay, dropped rates, etc.). Therefore, directly applying LTE scheduling algorithms for M2M traffic will results in sub-optimal performance. Hence, there is a need for new scheduling algorithms. It should be noted that dynamic scheduling policies would lead to unacceptable signalling load due to a large number of devices. Thus, static scheduling policies are required that grant access to M2M devices based on periodic traffic patterns.

8.3.2 Overload Control and Signalling Reduction Techniques

It is expected that a multitude of cellular connected M2M devices will exist in near future. Hence, there is a need to develop overload control mechanisms, especially in scenarios when a large number of devices request resources at the same time. Existing solutions based

on Access Class Barring (ACB) have been shown to be sub-optimal [171]. Moreover, ACB based solutions require additional overhead in terms of broadcasting the barring information. An efficient approach to handle congestion is the introduction of congestion aware admission control mechanisms in the access network. For this purpose, various insights from control theory or the theory of autonomic networking can be obtained. Since most M2M devices generate small data transmissions, the signalling associated with each transmission should be minimized as well.

8.3.3 D2D-assisted M2M Communications

Device-to-Device (D2D) communications [172] for LTE-Advanced is currently under active investigation in academia, industry, and standardization bodies. Research shows that D2D communications has advantages of offloading traffic from cellular network, increased spectral efficiency, and reduced communication delay. However D2D creates new challenges in terms of interference management, resource allocation, device discovery, etc. Existing studies have mainly investigated D2D for H2H communications. An interesting avenue of future research is the application of D2D communications for Peer-to-Peer (P2P) communication between M2M devices.

8.4 Future Work for Nano-M2M

8.4.1 Routing in Diffusion-based Molecular Nano-M2M Networks

In literature, routing for diffusion-based molecular nano-M2M networks has been rarely investigated. Since the molecules propagate reliably over limited ranges, molecular communication will be realized mainly in a multi-hop manner. Therefore, routing functionalities becomes particularly important. A fundamental challenge in this regard is the addressing of nanomachines (NMs) as without specifying the receiver NM, the selection of next hop becomes challenging. Thus, there is a need to develop routing techniques with minimal addressing requirements and control overhead. A simple routing technique

is random flooding with probabilistic forwarding, where the probability of forwarding at each hop is a function of the concentration gradient of molecules.

8.5 Other Research Directions

8.5.1 Convergence of Capillary M2M and Cellular Networks

Recently, it has been proposed to converge cellular networks and wireless sensor networks for enhanced performance of both types of networks [173]. This concept can be extended for the convergence of capillary M2M networks and cellular networks. Conventionally, the capillary M2M network connected to a cellular network through a dual mode M2M gateway results in a hierarchical architecture. However, a convergence of both networks results in a flat architecture resulting in reduced hierarchical signalling. Moreover, it will provide higher layer control and optimization for the capillary M2M network through the cellular network. Key challenges for realizing this converged architecture include: (a) timing coordination between two networks as both networks employ different types of multiple access schemes, (b) air-interface convergence of the cellular interface and the capillary M2M interface, and (c) protocol stack convergence.

Appendices

Appendix A

Proof of Theorem 3.3.1

It should be noted that both B_i and n_t are random variables. Therefore, the average access delay can be calculated as follows.

$$E\left(w\right) = \mathbb{E}\left(\sum_{i=1}^{n_t} B_i\right) = \mathbb{E}_{n_t}\left(\mathbb{E}\sum_{i=1}^{n_t} B_i \;\middle|\; n_t\right), \tag{A.1}$$

where $\mathbb{E}_{n_t}\left(\cdot\right)$ denotes the expected value over the random variable n_t. Since $\mathbb{E}\left(B_i\right) = \left(1+\lambda_i\right)/2$,

$$E\left(w\right) = \mathbb{E}_{n_t}\left[\sum_{i=1}^{n_t}\frac{1+\lambda_i}{2}\right] = \frac{1}{2}\mathbb{E}_{n_t}\left[\sum_{i=1}^{n_t}\left(2^{\frac{\log\left(i\cdot\frac{\psi_m}{T_s}\right)}{\delta}}+1\right)\right]$$

For simplicity we assumed $\omega = 1$ in (3.1). Next, for the ease of analysis, we assume $\delta = \log 2$. Converting *Briggsian* logarithmic function to *Binary* logarithmic function, we obtain,

$$E\left(w\right) = \frac{1}{2}\mathbb{E}_{n_t}\left[\frac{\psi_m}{T_s}\frac{n_t(n_t+1)}{2}+1\right] = \frac{1}{4}\frac{\psi_m}{T_s}\mathbb{E}_{n_t}\left(n_t^2+n_t\right)+\frac{1}{2}\mathbb{E}_{n_t}\left(n_t\right)$$

We know that $\sum_{n=0}^{\infty} n^2 q^n = \frac{q(1+q)}{(1-q)^3}$; therefore, using (3.9), $N_t^{'}$ is calculated as follows.

$$N_t^{'} = \mathbb{E}\left(n_t^2\right) = \sum_{n=0}^{\infty} n_t^2\left(1-P_s\right)^n P_s = \frac{(1-P_s)(2-P_s)}{P_s^2}$$

Hence, the average access delay is given by

$$E\left(w\right) = \frac{1}{4}\frac{\psi_m}{T_s}\left(N_t^{'}+N_t\right)+\frac{1}{2}N_t \tag{A.2}$$

Note that the analysis is based on the assumption of no maximum retry limit for the MTCD in case of repeated collisions. If such a limit exists, the analysis will slightly change

whereby the probability of dropping a packet (P_{drop}) needs to be computed. Using P_{drop}, the average access delay can be calculated as given by (A.3), which is straightforward to obtain using the backoff model.

$$E(w) = \mathbb{E}\left[\sum_{i=1}^{n_t} B_i \;\middle|\; (1 - P_{drop})\right] \tag{A.3}$$

Appendix B

Proof of Equation (3.22)

Although the proof can be obtained from [106], it is given for the completeness of the chapter. Recall that the optimum tradeoff is achieved with an equality relationship i.e., $P_f = P_m$. Therefore, using (3.3) and (3.4),

$$\frac{1}{2}erfc\left(\frac{1}{\sqrt{2}}\frac{\sigma-2n}{\sqrt{4n}}\right) = 1 - \frac{1}{2}erfc\left(\frac{1}{\sqrt{2}}\frac{\sigma-2n\left(\gamma+1\right)}{\sqrt{4n\left(2\gamma+1\right)}}\right) \tag{B.1}$$

Note that $erfc(x) = 1 - erf(x)$, and $erf(x) < 0$ for $x < 0$. Hence, for the above relationship to exist,

$$\frac{1}{\sqrt{2}}\frac{\sigma-2n}{\sqrt{4n}} > 0 \quad \text{and} \quad \frac{1}{\sqrt{2}}\frac{\sigma-2n\left(\gamma+1\right)}{\sqrt{4n\left(2\gamma+1\right)}} < 0$$

It should be noted that $n = T_dW$ represents the time-bandwidth (channel detection time and channel bandwidth) product. Using the above inequalities, we find the optimal channel detection time from the following relationship.

$$\frac{1}{2}\left[\frac{1}{\sqrt{2}}\frac{\sigma-2n}{\sqrt{4n}} + \frac{1}{\sqrt{2}}\frac{\sigma-2n\left(\gamma+1\right)}{\sqrt{4n\left(2\gamma+1\right)}}\right] = 0 \Rightarrow T_d^* = \frac{\sigma}{2W\left[1+\left(\frac{\gamma}{1+\sqrt{2\gamma+1}}\right)\right]} \tag{B.2}$$

Alternatively, the optimal channel detection time can also be calculated in terms of P_d, by expressing P_f in terms of P_d and using the equality relationship. We only give the final expression here. For more details see [101] and [106].

$$T_d^* = \frac{2}{\gamma^2 W}\left[erfc^{-1}(2-2P_d) - erfc^{-1}(2P_d)\sqrt{2\gamma+1}\right]^2 \tag{B.3}$$

Appendix C

Proof of Theorem 3.4.1

It should be noted that $P_f = \frac{1}{2} erfc\left(\rho + \sqrt{\frac{T_d W}{2}}\gamma\right)$ [98], where $\rho = erfc^{-1}(2P_d^{'})\sqrt{2\gamma + 1}$. Therefore, (3.24) can be expressed as follows.

$$Y\left(T_d, T_{MF}\right) = \frac{1}{T_{MF}}\left(T_{MF} - T_d - T^{'}\right) \cdot \theta_{PRMA} \cdot P_i \cdot R, \tag{C.1}$$

where $R = \left[1 - \frac{1}{2} erfc\left(\rho + \sqrt{\frac{T_d W}{2}}\gamma\right) + \alpha\right]$ such that $\alpha = P_b(1 - P_d^{'})P_i^{-1}$.

Since $\frac{\partial}{\partial x} erf(x) = \frac{2}{\sqrt{\pi}}\exp(-x^2)$, we calculate the derivative of $Y\left(T_d, T_{MF}\right)$ w.r.t T_d as follows.

$$Y^{'}\left(T_d, T_{MF}\right) = \frac{\theta_{PRMA} \cdot P_i}{T_{MF}}\left[A^{'} + \left(T_{MF} - T_d - T^{'}\right) F^{'}\right]$$

where

$$A^{'} = \frac{1}{2} erfc\left(\rho + \sqrt{\frac{T_d W}{2}}\gamma\right) - 1 - \alpha,$$

and

$$F^{'} = \frac{\gamma}{2}\exp\left(-\left(\rho + \sqrt{\frac{T_d W}{2}}\gamma\right)^2\right)\sqrt{\frac{W}{2\pi T_d}}$$

Under the equality relationship i.e., $P_f = P_m = (1 - P_d^{'})$, and setting $Y^{'}\left(T_d, T_{MF}\right) = 0$, we obtain,

$$X_d^* = \frac{2}{\gamma T_F}\exp\left(\rho + \sqrt{\frac{W T_d^*}{2}}\gamma\right)^2 \sqrt{\frac{2\pi T_d^*}{W}}\left(P_d^{'} + \alpha\right) + \frac{T_{sw}}{T_F} \tag{C.2}$$

Appendix D

Proof of Theorem 3.4.2

According to the interference constraint, $IR \leq IR_{max}$. Therefore, using (3.27)

$$P_m P_b + P_i (1 - P_f) + e^{-\mu T} (P_f - P_d) \leq IR_{max} \tag{D.1}$$

Since $P_m = 1 - P_d^{'}$ (under detection constraint) and $T = T_p + X \cdot T_F$, we obtain the optimal reservation cycle as follows.

$$X_i^* = \frac{1}{\mu T_F} \left\{ \ln P_i - \mu T_p + \ln(1 - 2P_d^{'}) - \ln \left(IR_{max} + 2P_b P_d^{'} - (P_b + P_d^{'}) \right) \right\} \tag{D.2}$$

where $\ln(\cdot)$ denotes the natural logarithm. While X_i^* satisfies the interference constraint, it cannot guarantee that the throughput is maximized. In order to ensure throughput maximization under both detection and interference constraints, we evaluate X_d^* (from Theorem 3.4.1) in terms of X_i^*, by introducing P_d^* (given by (D.3)) and obtain the optimal reservation cycle, X^* in terms of IR_{max} and $P_d^{'}$ as given by (3.28).

$$P_d^* = \frac{1 - B(IR_{max} + P_b)}{2 + (2P_b - 1)B}, \tag{D.3}$$

where

$$B = P_i^{-1} \exp \left[\mu \left(X_i^* \cdot T_f + T_p \right) \right]$$

Appendix E

Proof of Theorem 3.4.3

According to the duty cycle constraint, together with interference and detection constraints, $DC^{'} \leq DC_{max}$. Therefore, using (3.17)

$$\begin{aligned} &\frac{1}{T_{MF}}\left[T_p + N_t T_s + (X-\beta)T_s + \chi_{DL}T_s + P_{cm}T_s\right] \leq DC_{max} \\ &\Rightarrow N_t \leq \left[DC_{max}T_{MF}^{'} - T_p\right] T_s^{-1} - X^{'} + \beta - \chi_{DL} \end{aligned} \tag{E.1}$$

Let, $K^{'} = \left[DC_{max}T_{MF}^{'} - T_p\right] T_s^{-1}$. Therefore, $N_t \leq K^{'} - X^{'} + \beta - \chi_{DL}$. It should be noted that $P_{cm} \ll 1$. Hence, we can ignore its effect in the expression for duty cycle. Recall that N_t gives the expected number of transmissions required to obtain reservation.

We obtain the Cumulative Distribution Function (CDF) of n_t , denoted by $F(x)$, using the PMF in (3.9) as follows.

$$F(x) \triangleq Pr\left\{n_t \leq x\right\} = \sum_{n=0}^{x} P_n = \sum_{n=0}^{x} (1-P_s)^n P_s$$

Note that $\sum_{k=0}^{n-1} r^k = \frac{1-r^n}{1-r}$; therefore,

$$F(x) = \frac{P_s\left[1-(1-P_s)^{x+1}\right]}{1-P_s}$$

Using $F(x)$, we obtain the probability of violating the duty cycle constraint as follows.

$$Pr\left\{DC^{'} > DC_{max}\right\} = Pr\left\{n_t > K^{'} - X^{'} + \beta - \chi_{DL}\right\} = 1 - \frac{P_s\left[1-(1-P_s)^{\mathcal{Z}}\right]}{1-P_s}, \tag{E.2}$$

where $\mathcal{Z} = K^{'} - X^{'} + \beta - \chi_{DL} + 1$.

Appendix F

Preliminaries on Statistical QoS Guarantees and Effective Capacity

In literature, statistical QoS guarantees have been extensively investigated in the context of *effective bandwidth* theory [174], [175]. The effective bandwidth is defined as the minimum constant service rate for a given arrival process, such that a required QoS constraint θ can be guaranteed. The QoS constraint, θ characterizes the queue length decaying rate. Specifically, for a dynamic queuing system, under sufficient conditions, the queue length process, $\mathcal{Q}(t)$ converges in distribution to a random variable $\mathcal{Q}(\infty)$ such that

$$-\lim_{x_t \to 0} \frac{\ln\left(Pr\left\{\mathcal{Q}(\infty) > x_t\right\}\right)}{x_t} = \theta \tag{F.1}$$

The above equation states that the probability of the queue length exceeding a certain threshold x_t decays exponentially fast as x_t increases and the parameter θ determines the decaying rate.

Motivated by the effective bandwidth theory, Wu and Negi proposed a dual concept, termed as *effective capacity* [143]. The effective capacity is defined as the maximum constant arrival rate that a given service process can support in order to guarantee a QoS requirement specified by θ.

Let, $R[i]$ denote a discrete time stationary and ergodic service process and $S[t] = \sum_{i=1}^{t} R[i]$ be the corresponding time accumulated process, where $R[i]$ represents the amount of data served in the i^{th} time frame. We assume that the asymptotic log-moment generating function of $S[t]$, defined as

$$\aleph(\theta) \triangleq \lim_{t \to \infty} \frac{1}{t} \ln \mathbb{E}\left(e^{\theta S[t]}\right),$$

exists for all $\theta \geq 0$. Under this condition, the effective capacity of such a service process is given as follows [143].

$$C_E(\theta) \triangleq -\frac{\aleph(-\theta)}{\theta} = -\lim_{t\to\infty}\frac{1}{\theta t}\ln \mathbb{E}\left(e^{-\theta S[t]}\right) \tag{F.2}$$

When the service process, $R[i]$ is uncorrelated, the effective capacity simplifies to

$$C_E(\theta) = -\frac{1}{\theta}\ln \mathbb{E}\left(e^{-\theta R[i]}\right) \tag{F.3}$$

Appendix G

Calculating $\mathbb{E}\left(e^{-\theta_k R_k}\right)$

Recall that the SNR for the k^{th} user over the l^{th} RB is given by $\gamma_{k,l} = \frac{P_{k,l}|h_{k,l}|^2}{\sigma^2}$. The channel coefficient, $h_{k,l}$ is given by $h_{k,l} = F_{k,l}\sqrt{\frac{1}{Z_k}}$, where $F_{k,l}$ represents the channel fading coefficient and Z_k is the path loss. The channel fading coefficient, $F_{k,l}$, is modeled as independent circularly symmetric complex Gaussian distributed random variable with zero mean and unit variance. The expected value of the channel power gain is given by $\mathbb{E}\left(\mid h_{k,l} \mid^2\right) = \sigma^2_{k,l} = \frac{1}{\xi D_k^{\alpha}}$, where ξ depends on the choice of path loss model, D_k is the distance between the user k and the eNB, and α is the path loss exponent. It follows that $\mid h_{k,l} \mid^2$ is exponentially distributed and the PDF is given by $\frac{1}{\sigma^2_{k,l}} \exp\left(-\frac{|h_{k,l}|^2}{\sigma^2_{k,l}}\right)$ for $\mid h_{k,l} \mid^2 \geq 0$.

It should be noted that if x is an exponential random variable with rate parameter r, then $c \cdot x$ is exponentially distributed with rate parameter r/c, where $c > 0$ is a constant [165]. Therefore, the PDF of $\gamma_{k,l}$ is given by $\frac{\sigma^2}{P_{k,l}\sigma^2_{k,l}} \exp\left(-\frac{\sigma^2|h_{k,l}|^2}{P_{k,l}\sigma^2_{k,l}}\right)$.

Using (6.1), $\mathbb{E}\left(e^{-\theta_k R_k}\right) = \mathbb{E}\left(\{1 + \mu \cdot \gamma_{eff,k}\}^t\right)$, where $t = \frac{-\theta_k B L_k}{\ln 2}$. In order to calculate this expected value, we need the PDF of $\gamma_{eff,k}$, which is obtained as follows.

We assume that the channel on each RB allocated to a user is independent and identically distributed (i.i.d).

Note that *the sum of Q mutually independent Exponential random variables with parameter χ is Gamma distributed with parameters $2Q$ and Q/χ* [165].

This is proved using the moment generating functions (MGFs). Let $Z = \sum_{i=1}^{Q} Y_i$, where Y_i is exponentially distributed with parameter χ. The MGF of a sum of mutually independent random variables is equal to the product of their MGFs. Hence,

$$M_z(t) = \prod_{i=1}^{Q} M_{y_i}(t) = \prod_{i=1}^{Q} \frac{\chi}{\chi - t} = \left(1 - \frac{2(\frac{Q}{\chi})}{2Q}t\right)^{\frac{-2Q}{2}}, \tag{G.1}$$

which is the MGF of Gamma distribution with parameters $2Q$ and Q/χ. Therefore, Z is Gamma distributed as two random variables have the same distribution when they have the same MGF.

From (6.2) it can be seen that $\gamma_{eff,k}$ is a sum of exponential random variables. Thus, $\gamma_{eff,k}$ is Gamma distributed with parameters $\alpha' = 2L_k$ and $\beta' = P_{k,l}\sigma_{k,l}^2/\sigma^2$, and the PDF is given by

$$f(\gamma_{eff,k}, \alpha', \beta') = \begin{cases} \frac{{\gamma_{eff,k}}^{\alpha'-1}}{\Gamma(\alpha')(\beta')^{\alpha'}} \exp\left(-\frac{\gamma_{eff,k}}{\beta'}\right) & \gamma_{eff,k} \geq 0 \\ 0 & \gamma_{eff,k} < 0 \end{cases} \tag{G.2}$$

where $\Gamma(\alpha') = \int_0^\infty {\gamma_{eff,k}}^{\alpha'-1} \exp\Big(-\gamma_{eff,k}\Big)\partial\gamma_{eff,k}$.

Define $U = g\left(\gamma_{eff,k}\right) = \left\{1 + \mu.\gamma_{eff,k}\right\}^t$ and denote its PDF by $f_U(U)$, which is given by $\frac{f(\gamma^*_{eff,k},\alpha',\beta')}{g'\left(\gamma^*_{eff,k}\right)}$ [165], where $\gamma^*_{eff,k}$ denotes the root of $U = g(\gamma_{eff,k})$, and $g'(\gamma^*_{eff,k})$ denotes the derivative of $g(\gamma_{eff,k})$ evaluated at $\gamma^*_{eff,k}$. Thus $f_U(U)$ is given by

$$f_U\left(U\right) = \frac{f(\frac{U^{1/t}-1}{\mu}, \alpha', \beta')}{\mu t (U^{1/t})^{t-1}}, \tag{G.3}$$

and therefore,

$$\mathbb{E}\left(\left\{1 + \mu \cdot \gamma_{eff,k}\right\}^t\right) = \mathbb{E}\left(U\right) = \int_0^\infty U f_U\left(U\right) \partial U \tag{G.4}$$

Appendix H

Proof of Lemma 6.2.1

The optimal power control policy ($\mu_{opt,k}$), that maximizes the effective capacity for a user k is derived through the following optimization problem [146].

$$\max\left\{-\frac{1}{\theta_k}\ln\mathbb{E}\left(e^{-\theta_k R_k}\right)\right\} = \min\left\{\frac{1}{\theta_k}\ln\mathbb{E}\left(e^{-\theta_k R_k}\right)\right\} \tag{H.1}$$
$$s.t \quad \int_0^\infty \mu f(\gamma_{eff,k},\alpha',\beta')\partial_{\gamma_{eff,k}} = 1$$

Using (6.1), $\mathbb{E}\left(e^{-\theta_k R_k}\right) = \mathbb{E}\left(\{1+\mu\cdot\gamma_{eff,k}\}^t\right)$, where $t = \frac{-\theta_k BL_k}{\ln 2}$. The Lagrangian function of the optimization problem (H.1) is given by

$$\check{j} = \int_0^\infty (1+\mu\cdot\gamma_{eff,k})^t f(\gamma_{eff,k},\alpha',\beta')\partial_{\gamma_{eff,k}} + \lambda\Big(\int_0^\infty \mu f(\gamma_{eff,k},\alpha',\beta')\partial_{\gamma_{eff,k}} - 1\Big) \tag{H.2}$$

where $f(\gamma_{eff,k},\alpha',\beta')$ is given by (G.2). Taking the derivative of the Lagrangian function and equating to zero, we get

$$\frac{\partial\check{j}}{\partial\mu} = \left\{t(1+\mu\cdot\gamma_{eff,k})^{t-1}\gamma_{eff,k}+\lambda\right\} f(\gamma_{eff,k},\alpha',\beta') = 0$$

Setting $\gamma_0 = -\frac{\lambda}{t}$, we obtain

$$\gamma_0 = \gamma_{eff,k}(1+\mu\cdot\gamma_{eff,k})^{t-1} \Rightarrow \mu = \mu_{opt,k} = \frac{1}{\gamma_{eff,k}}\left[\left(\frac{\gamma_0}{\gamma_{eff,k}}\right)^{\frac{1}{t-1}} - 1\right]$$

The cutoff effective SNR, γ_0 is obtained by numerically solving (H.3).

$$\int_{\gamma_0}^\infty \frac{1}{\gamma_{eff,k}}\left[\left(\frac{\gamma_0}{\gamma_{eff,k}}\right)^{\frac{1}{t-1}} - 1\right] f(\gamma_{eff,k},\alpha',\beta')\partial_{\gamma_{eff,k}} = 1 \tag{H.3}$$

Appendix I

Proof of Theorem 6.4.1

The proof of dual relationship between the canonical dual problem and the primal problem in (6.7) directly extends from [153].

The canonical dual problem is perfectly dual to the primal problem if $\mathcal{C}_j(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}) = \Omega(\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*)$, where $\widetilde{\boldsymbol{x}}$ denotes the KKT point of primal problem for a given $\boldsymbol{\theta}$ and $\widetilde{\mathbf{y}}^* = \Omega(\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*) \in \mathcal{X}_p^\sharp$ denotes the KKT point of the dual function. This is proved using (6.12). The total complementarity function at the KKT point $(\widetilde{\boldsymbol{x}}, \widetilde{\mathbf{y}}^*)$ becomes

$$\Xi(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}, \widetilde{\mathbf{y}}^*) = \wedge(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}})^T \widetilde{\mathbf{y}}^* - V^*(\widetilde{\mathbf{y}}^*) + \mathcal{C}_j(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}) \tag{I.1}$$

Let, $\boldsymbol{\lambda}_{\boldsymbol{\delta}}, \boldsymbol{\lambda}_{\boldsymbol{\beta}}, \boldsymbol{\lambda}_{\boldsymbol{\tau}}$, and $\boldsymbol{\lambda}_{\boldsymbol{\sigma}}$ be the Lagrange multipliers associated with the inequality $\widetilde{\mathbf{y}}^* = (\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*) > 0$. According to the complementarity condition: $(\boldsymbol{\lambda}_{\boldsymbol{\delta}}, \boldsymbol{\lambda}_{\boldsymbol{\beta}}, \boldsymbol{\lambda}_{\boldsymbol{\tau}}, \boldsymbol{\lambda}_{\boldsymbol{\sigma}}) \leq 0$, $(\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*) > 0$, $(\widetilde{\boldsymbol{\delta}}^*)^T \boldsymbol{\lambda}_{\boldsymbol{\delta}} = 0$, $(\widetilde{\boldsymbol{\beta}}^*)^T \boldsymbol{\lambda}_{\boldsymbol{\beta}} = 0$, $(\widetilde{\boldsymbol{\tau}}^*)^T \boldsymbol{\lambda}_{\boldsymbol{\tau}} = 0$, $(\widetilde{\boldsymbol{\sigma}}^*)^T \boldsymbol{\lambda}_{\boldsymbol{\sigma}} = 0$. Thus, $\wedge(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}})^T = 0$. Moreover, $V^*(\widetilde{\mathbf{y}}^*) = 0$ for $\widetilde{\mathbf{y}}^* > 0$. Therefore, $\Xi(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}, \widetilde{\mathbf{y}}^*) = \mathcal{C}_j(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}})$. This completes the first part of proof.

Using (6.13), $\Xi(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}, \widetilde{\mathbf{y}}^*) = \Xi(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*)$ is given by

$$\begin{aligned}\Xi(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*) = &\sum_{k=1}^{K}\sum_{a=1}^{A} \widetilde{x}_{k,a}\Big(\widetilde{\sigma}_k^* - \widetilde{\beta}_{k,a}^* + \sum_{l=1}^{L} \widetilde{\tau}_l^* M_{l,a}^k\Big) + \frac{\sum_{k=1}^{K} \frac{1}{\theta_k} \ln \mathbb{E}\left(e^{-\theta_k R_k}\right)}{\sum_{k=1}^{K}\sum_{a=1}^{A} \widetilde{x}_{k,a} P_{k,a}} - \sum_{k=1}^{K} \widetilde{\sigma}_k^* \\ &- \sum_{l=1}^{L} \widetilde{\tau}_l^* + \sum_{k=1}^{K} \widetilde{\delta_k}^* \mathbb{E}\left(e^{-\theta_k R_k}\right) - \sum_{k=1}^{K} \widetilde{\delta_k}^* e^{-\theta_k \mathcal{B}_k} + \sum_{k=1}^{K}\sum_{a=1}^{A} \widetilde{\beta}_{k,a}^* (\widetilde{x}_{k,a})^2\end{aligned} \tag{I.2}$$

Since $\widetilde{\boldsymbol{x}}$ denotes the KKT point of the primal problem, which is obtained by $\nabla_{\widetilde{x}} \Xi(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*) = 0$,

$$\widetilde{x}_{k,a}(\theta_k, \widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*) = \frac{\ln \mathbb{E}\left(e^{-\theta_k R_k}\right)}{2\theta_k P_{k,a} \widetilde{\beta}_{k,a}^*} - 0.5 \frac{\widetilde{w}}{\widetilde{\beta}_{k,a}^*}$$

where $\widetilde{w} = \left(\widetilde{\sigma}_k^* - \widetilde{\beta}_{k,a}^* + \sum_{l=1}^{L} \widetilde{\tau}_l^* M_{l,a}^k\right)$. Substituting for $\widetilde{x}_{k,a}$ in (I.2) we obtain

$$\begin{aligned}\Xi(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*) = \frac{\sum_{k=1}^{K} \frac{1}{\theta_k} \ln \mathbb{E}\left(e^{-\theta_k R_k}\right)}{\sum_{k=1}^{K} \sum_{a=1}^{A} \frac{\widetilde{\Psi}}{\widetilde{\beta}_{k,a}^*}} + \sum_{k=1}^{K} \widetilde{\delta_k}^* \mathbb{E}\left(e^{-\theta_k R_k}\right) \\ - \sum_{k=1}^{K} \widetilde{\delta_k}^* e^{-\theta_k \mathcal{B}_k} - \sum_{k=1}^{K} \widetilde{\sigma}_k^* - \sum_{l=1}^{L} \widetilde{\tau_l}^* + \sum_{k=1}^{K} \sum_{a=1}^{A} \frac{\widetilde{\Psi}}{2P_{k,a}\widetilde{\beta}_{k,a}^*} \Big(\frac{\widetilde{\Psi}}{2P_{k,a}} + \widetilde{w} \Big)\end{aligned} \tag{I.3}$$

$= \Omega(\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*)$, where $\widetilde{\Psi} = \frac{1}{\theta_k} \ln \mathbb{E}\left(e^{-\theta_k R_k}\right) - P_{k,a}\widetilde{w}$. This completes the proof.

Appendix J

Proof of Theorem 6.4.2

The total complementarity function, $\Xi(\boldsymbol{\theta}, \boldsymbol{x}, \mathbf{y}^*)$ is convex in $(\boldsymbol{\theta}, \boldsymbol{x})$ and concave in $\mathbf{y}^*$. Therefore, the stationary point (for a given $\boldsymbol{\theta}$), $(\widetilde{\boldsymbol{x}}, \widetilde{\mathbf{y}}^*)$ is a saddle point of $\Xi(\boldsymbol{\theta}, \boldsymbol{x}, \mathbf{y}^*)$ [153].

The dual function $\Omega(\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*)$ is concave on $\mathcal{X}_p^\sharp$ and the KKT point $\widetilde{\mathbf{y}}^* = (\widetilde{\boldsymbol{\delta}}^*, \widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\tau}}^*, \widetilde{\boldsymbol{\sigma}}^*)$ must be its global maximizer. Thus, by saddle min-max duality theory, we have

$$\begin{aligned}
\Omega(\widetilde{\mathbf{y}}^*) &= \max_{\mathbf{y}^* \in \mathcal{X}_p^\sharp} \Omega(\mathbf{y}^*) = \max_{\mathbf{y}^* \in \mathcal{X}_p^\sharp} \min_{(\boldsymbol{\theta}, \boldsymbol{x}) \in \mathcal{X}_p} \Xi(\boldsymbol{\theta}, \boldsymbol{x}, \mathbf{y}^*) \\
&= \min_{(\boldsymbol{\theta}, \boldsymbol{x}) \in \mathcal{X}_p} \max_{\mathbf{y}^* \in \mathcal{X}_p^\sharp} \Xi(\boldsymbol{\theta}, \boldsymbol{x}, \mathbf{y}^*) \\
&= \min_{(\boldsymbol{\theta}, \boldsymbol{x}) \in \mathcal{X}_p} \left\{ \mathcal{C}_j(\boldsymbol{\theta}, \boldsymbol{x}) + \max_{\mathbf{y}^* \in \mathcal{X}_p^\sharp} \wedge(\boldsymbol{\theta}, \boldsymbol{x})^T \mathbf{y}^* - V^*(\mathbf{y}^*) \right\} \\
&= \min_{(\boldsymbol{\theta}, \boldsymbol{x}) \in \mathcal{X}_p} \mathcal{C}_j(\boldsymbol{\theta}, \boldsymbol{x}) = C_j(\boldsymbol{\theta}, \widetilde{\boldsymbol{x}})
\end{aligned}$$

due to the fact that

$$V(\wedge(\boldsymbol{\theta}, \boldsymbol{x})) = \sup_{\mathbf{y}^* \in \mathcal{X}_p^\sharp} \left\{ \wedge(\boldsymbol{\theta}, \boldsymbol{x})^T \mathbf{y}^* - V^\star(\mathbf{y}^*) \right\}$$

$$= \begin{cases} 0 & \text{if } (\boldsymbol{\theta}, \boldsymbol{x}) \in \mathcal{X}_p \\ +\infty & \text{otherwise} \end{cases}$$

This completes the proof.

Appendix K

Analytical Modeling for CSB-MAC

The expressions for CSB-MAC protocol can be obtained using a similar framework as described for CRB-MAC in Section 4.3.

The single hop energy consumption for CSB-MAC is given by $\mathcal{E}^j_{CSB\_total} = \chi_{CSB}(\mathcal{E}\_T^j_{fail} + \mathcal{E}\_R^j_{fail}) + \mathcal{E}\_T^j_{succ} + \mathcal{E}\_R^j_{succ} + \chi^j_{ss}\mathcal{E}^j_{ss}$. The number of retransmissions, χ_{CSB} can be calculated in a similar way as described in (4.15), with $P_a = (P^j_{fail})^a(1 - P^j_{fail})$.

The single hop delay for CSB-MAC is given by $\mathcal{D}^j_{CSB} = \chi_{CSB} \cdot (T_{pr} + T_d + T_{CW}) + \chi^j_{ss} \cdot T_s + T_{CI} \cdot (\chi^j_{ss} - 1)$.

Finally, the single-hop reliability of CSB-MAC is given by $\mathcal{R}^j_{CSB} = 100 \times \left(1 - (P^j_{fail})^{(\chi_{CSB}+1)}\right)$.

Appendix L

References

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A Survey," *J. Comput. Networks*, vol. 54, no. 15, pp. 2787 – 2805, 2010.

[2] X. Li, R. Lu, X. Liang, X. Shen, J. Chen, and X. Lin, "Smart Community: An Internet of Things Application," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 68–75, 2011.

[3] G. Wu, S. Talwar, K. Johnsson, N. Himayat, and K. Johnson, "M2M: From Mobile to Embedded Internet," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 36–43, 2011.

[4] D. Boswarthick, O. Elloumi, and O. Hersent, *M2M Communications: A Systems Approach*. John Wiley, 2012.

[5] K. Chang, A. Soong, M. Tseng, and Z. Xiang, "Global Wireless Machine-to-Machine Standardization," *IEEE Internet Comput.*, vol. 15, no. 2, pp. 64–69, 2011.

[6] S. Haykin, "Cognitive Radio: Brain-empowered Wireless Communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb 2005.

[7] Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie, and S. Gjessing, "Cognitive Machine-to-Machine Communications: Visions and Potentials for the Smart Grid," *IEEE Network*, vol. 26, no. 3, pp. 6–13, 2012.

[8] S. Amin and B. Wollenberg, "Toward a Smart Grid: Power Delivery for the 21st Century," *IEEE Power and Energy Mag.*, vol. 3, no. 5, pp. 34–41, 2005.

[9] A. El-Hoiydi, "Aloha with Preamble Sampling for Sporadic Traffic in Ad-hoc Wireless Sensor Networks," in *IEEE International Conference on Communications (ICC)*, vol. 5, May 2002, pp. 3418–3423 vol.5.

[10] T. Winter, "IPv6 Routing Protocol for Low Power and Lossy Networks," Internet Engineering Task Force, RFC 6550, Mar. 2012. [Online]. Available: http://www.rfc-editor.org/rfc/rfc6550.txt

[11] D. Y. Gao, *Duality Principles in Non-Convex Systems: Theory, Methods and Applications*. Boston/Dordrecht/London: Kluwer, 2000.

[12] A. Mehrabian and C. Lucas, "A Novel Numerical Optimization Algorithm Inspired From Weed Colonization ," *J. Eco. Inform.*, vol. 1, no. 4, pp. 355 – 366, 2006.

[13] I. F. Akyildiz, F. Brunetti, and C. Blázquez, "Nanonetworks: A New Communication Paradigm," *J. Comput. Networks*, vol. 52, no. 12, pp. 2260–2279, Aug. 2008.

[14] I. Akyildiz and J. Jornet, "The Internet of Nano-Things," *IEEE Wireless Commun.*, vol. 17, no. 6, pp. 58–63, December 2010.

[15] T. Suda, M. Moore, T. Nakano, R. Egashira, and A. Enomoto, "Exploratory Research on Molecular Communication between Nanomachines," in *ACM Conference on Genetic and Evolutionary Computation (GECCO)*, 2005.

[16] I. Akyildiz, F. Fekri, R. Sivakumar, C. Forest, and B. Hammer, "MONACO:

Fundamentals of Molecular Nano-Communication Networks," *IEEE Wireless Commun.*, vol. 19, no. 5, pp. 12–18, Oct 2012.

[17] 3GPP, "Service Requirements for Machine Type Communications," 3rd Generation Partnership Project (3GPP), TS 22.368, Jun. 2010. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/22368.htm

[18] Ofcom, "Digital Dividend Review: Statement on License Exempted Devices using Interleaved Spectrum," http://stakeholders.ofcom.org.uk/binaries/consultations/cognitive/summary/cognitive.pdf, 2009, [Online; accessed Feb-2012].

[19] M. Nekovee, "Quantifying the Availability of TV White Spaces for Cognitive Radio Operation in the UK," in *IEEE International Conference on Communications (ICC) Workshops*, 2009, pp. 1–5.

[20] J. Palicot, "Cognitive Radio: An Enabling Technology for the Green Radio Communications Concept," in *ACM International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2009, pp. 489–494.

[21] M. Palattella, N. Accettura, X. Vilajosana, T. Watteyne, L. Grieco, G. Boggia, and M. Dohler, "Standardized Protocol Stack for the Internet of (Important) Things," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 3, pp. 1389–1406, Third Quarter 2013.

[22] IEEE 802.15.4 , "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)," http://www.ieee802.org/15, September 2006.

[23] J. Rudell, V. Bhagavatula, and W. Wesson, "Future Integrated Sensor Radios for Long-Haul Communication," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 101–109, April 2014.

[24] "IEEE Standard for Local and Metropolitan Area Networks–Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC Sublayer," *IEEE Std 802.15.4e-2012 (Amendment to IEEE Std 802.15.4-2011)*, pp. 1–225, April 2012.

[25] S. Tozlu, M. Senel, W. Mao, and A. Keshavarzian, "Wi-Fi Enabled Sensors for Internet of Things: A Practical Approach," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 134–143, June 2012.

[26] G. Montenegro, "Transmission of IPv6 Packets over IEEE 802.15.4 Networks," Internet Engineering Task Force, RFC 4944, Sep. 2007. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4944.txt

[27] J. Postel, "Transmission Control Protocol," Internet Engineering Task Force, RFC 793, Sep. 1981. [Online]. Available: http://www.rfc-editor.org/rfc/rfc793.txt

[28] ——, "User Datagram Protocol," Internet Engineering Task Force, RFC 768, Aug. 1980. [Online]. Available: http://www.rfc-editor.org/rfc/rfc768.txt

[29] Z. Shelby, "Constrained Application Protocol (CoAP)," Internet Engineering Task Force CoRE Working Group, Tech. Rep., Feb. 2011.

[30] "Constrained RESTful Environments (CoRE)," Internet Engineering Task Force CoRE Working Group, Tech. Rep. [Online]. Available: http://www.ietf.org/dyn/wg/charter/core-charter.html

[31] IEEE 802.15 working group, "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)," http://www.ieee802.org/15/pub/TG4m.html, [Online; accessed Jun-2012].

[32] C.-H. Lo and N. Ansari, "The Progressive Smart Grid System from Both Power and Communications Aspects," *IEEE Commun. Surveys & Tut.*, vol. 14, no. 3, pp. 799–821, 2012.

[33] K. C. Budka, J. G. Deshpande, T. L. Doumi, M. Madden, and T. Mew, "Communication Network Architecture and Design Principles for Smart Grids," *Bell Labs Technical Journal*, vol. 15, no. 2, pp. 205–227, 2010.

[34] H. Farhangi, "The Path of the Smart Grid," *IEEE Power and Energy Mag.*, vol. 8, no. 1, pp. 18–28, 2010.

[35] L. Tsoukalas and R. Gao, "From Smart Grids to an Energy Internet: Assumptions, Architectures and Requirements," in *International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, April 2008, pp. 94–98.

[36] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug 2002.

[37] ——, "Wireless Sensor Networks: A Survey ," *J. Comput. Networks*, vol. 38, no. 4, pp. 393 – 422, 2002.

[38] Y. Liu, C. Yuen, J. Chen, and X. Cao, "A Scalable Hybrid MAC Protocol for Massive M2M Networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2013, pp. 250–255.

[39] P. Sthapit and J.-Y. Pyun, "Passive Synchronization Based Energy-Efficient MAC Protocol over M2M Wireless Networks," *Intl. J. of Dist. Sensor Networks*, 2013.

[40] W. Xie, M. Goyal, H. Hosseini, J. Martocci, Y. Bashir, E. Baccelli, and A. Durresi, "Routing Loops in DAG-Based Low Power and Lossy Networks," in *IEEE International Conference on Advanced Information Networking and Applications (AINA)*, April 2010, pp. 888–895.

[41] ——, "A Performance Analysis of Point-to-Point Routing along a Directed Acyclic Graph in Low Power and Lossy Networks," in *IEEE International Conference on Network-Based Information Systems (NBiS)*, Sept 2010, pp. 111–116.

[42] J. Tripathi, J. De Oliveira, and J. P. Vasseur, "A Performance Evaluation Study of RPL: Routing Protocol for Low power and Lossy Networks," in *Annual Conference on Information Sciences and Systems (CISS)*, March 2010, pp. 1–6.

[43] N. Accettura, L. Grieco, G. Boggia, and P. Camarda, "Performance analysis of the RPL Routing Protocol," in *IEEE International Conference on Mechatronics (ICM)*, April 2011, pp. 767–772.

[44] O. Gaddour, A. Koubaa, S. Chaudhry, M. Tezeghdanti, R. Chaari, and M. Abid, "Simulation and Performance Evaluation of DAG construction with RPL," in *International Conference on Communications and Networking (ComNet)*, March 2012, pp. 1–8.

[45] N. Tsiftes, J. Eriksson, N. Finne, F. Österlind, J. Höglund, and A. Dunkels, "A Framework for Low-power IPv6 Routing Simulation, Experimentation, and Evaluation," *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 479–480, Aug. 2010.

[46] J. Ko, S. Dawson-Haggerty, O. Gnawali, D. Culler, and A. Terzis, "Evaluating the Performance of RPL and 6LoWPAN in TinyOS," in *IEEE/ACM Workshop on Extending the Internet to Low power and Lossy Networks (IPSN)*, April 2011.

[47] K. Heurtefeux and H. Menouar, "Experimental Evaluation of a Routing Protocol for Wireless Sensor Networks: RPL under Study," in *IFIP Wireless and Mobile Networking Conference (WMNC)*, April 2013, pp. 1–4.

[48] O. Gaddour and A. Koubâa, "RPL in a Nutshell: A Survey," *J. Comput. Net.*, vol. 56, no. 14, pp. 3163 – 3178, 2012.

[49] B. Pavković, F. Theoleyre, and A. Duda, "Multipath Opportunistic RPL Routing over IEEE 802.15.4," in *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, 2011, pp. 179–186.

[50] T. Clausen and U. Herberg, "Multipoint-to-Point and Broadcast in RPL," in *International Conference on Network-Based Information Systems (NBiS)*, Sept 2010, pp. 493–498.

[51] L. Saad and B. Tourancheau, "Sinks Mobility Strategy in IPv6-Based WSNs for Network Lifetime Improvement," in *IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Feb 2011, pp. 1–5.

[52] A. Dvir, T. Holczer, and L. Buttyan, "VeRA - Version Number and Rank Authentication in RPL," in *IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, Oct 2011, pp. 709–714.

[53] A. Dunkels, J. Alonso, and T. Voigt, "Making TCP/IP Viable for Wireless Sensor Networks," in *European Wireless Sensor Networks Conference (EWSN)*, 2004.

[54] S. Raza, H. Shafagh, K. Hewage, R. Hummen, and T. Voigt, "Lithe: Lightweight Secure CoAP for the Internet of Things," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3711–3720, Oct 2013.

[55] T. Potsch, K. Kuladinithi, M. Becker, P. Trenkamp, and C. Goerg, "Performance Evaluation of CoAP Using RPL and LPL in TinyOS," in *International Conference on New Technologies, Mobility and Security (NTMS)*, May 2012, pp. 1–5.

[56] C. Cormio and K. R. Chowdhury, "A Survey on MAC Protocols for Cognitive Radio Networks," *J. Ad-Hoc Networks*, vol. 7, no. 7, pp. 1315 – 1329, 2009.

[57] A. Lioumpas and A. Alexiou, "Uplink Scheduling for Machine-to-Machine Communications in LTE-based Cellular systems," in *IEEE GLOBECOM Workshops (GC Wkshps)*, Dec 2011, pp. 353–357.

[58] J. Brown and J. Khan, "Predictive Resource Allocation in the LTE uplink for Event Based M2M Applications," in *IEEE International Conference on Communications (ICC) Workshops*, June 2013, pp. 95–100.

[59] S.-Y. Lien and K.-C. Chen, "Massive Access Management for QoS Guarantees in 3GPP Machine-to-Machine Communications," *IEEE Commun. Letters*, vol. 15, no. 3, pp. 311–313, March 2011.

[60] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward Ubiquitous Massive Accesses in 3GPP Machine-to-Machine Communications," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 66–74, 2011.

[61] A. G. Gotsis, A. S. Lioumpas, and A. Alexiou, "Analytical Modelling and Performance Evaluation of Realistic Time-Controlled M2M Scheduling over LTE Cellular Networks," *Trans. Emerging Tel. Tech.*, vol. 24, no. 4, pp. 378–388, 2013.

[62] K. Wang, J. Alonso-Zarate, and M. Dohler, "Energy-Efficiency of LTE for Small Data Machine-to-Machine Communications," in *IEEE International Conference on Communications (ICC)*, June 2013, pp. 4120–4124.

[63] I. M. Delgado-Luque, F. Blanquez-Casado, F. J. Martin-Vega, M. G. Fuertes, G. Gomez, M. C. Aguayo-Torres, J. Entrambasaguas, and J. Banos, "Performance Evaluation of Cooperation-based Techniques for M2M Traffic over LTE," in *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 144–148.

[64] 3GPP, "Access Class Barring and Overload Protection (ACBOP)," 3rd Generation Partnership Project (3GPP), TR 23.898, 2007. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/23898.htm

[65] S. Duan, V. Shah-Mansouri, and V. W. Wong, "Dynamic Access Class Barring for M2M Communications in LTE Networks," in *IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 4876–4881.

[66] A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "Cellular-based machine-to-machine: overload control," *IEEE Network*, vol. 26, no. 6, pp. 54–60, November 2012.

[67] A. Lo, Y. Law, and M. Jacobsson, "A Cellular-Centric Service Architecture for Machine-to-Machine (M2M) Communications," *IEEE Wireless Commun.*, vol. 20, no. 5, pp. 143–151, October 2013.

[68] M. Pierobon and I. Akyildiz, "A Physical End-to-End Model for Molecular Communication in Nanonetworks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 4, pp. 602–611, May 2010.

[69] I. F. Akyildiz and J. M. Jornet, "Electromagnetic Wireless Nanosensor Networks ," *J. Nano Commun. Networks*, vol. 1, no. 1, pp. 3 – 19, 2010.

[70] J. Jornet and I. Akyildiz, "Channel Modeling and Capacity Analysis for Electromagnetic Wireless Nanonetworks in the Terahertz Band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, October 2011.

[71] ——, "Graphene-based Plasmonic Nano-Antenna for Terahertz Band Communication in Nanonetworks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 685–694, December 2013.

[72] K. V. Srinivas, A. Eckford, and R. Adve, "Molecular Communication in Fluid Media: The Additive Inverse Gaussian Noise Channel," *IEEE Trans. Inform. Theory*, vol. 58, no. 7, pp. 4678–4692, July 2012.

[73] M. Kuran, H. Yilmaz, T. Tugcu, and I. Akyildiz, "Modulation Techniques for Communication via Diffusion in Nanonetworks," in *IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–5.

[74] H. Arjmandi, A. Gohari, M. Kenari, and F. Bateni, "Diffusion-Based Nanonetworking: A New Modulation Technique and Performance Analysis," *IEEE Commun. Letters*, vol. 17, no. 4, pp. 645–648, April 2013.

[75] T. Nakano, Y. Okaie, and A. Vasilakos, "Transmission Rate Control for Molecular Communication among Biological Nanomachines," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 835–846, December 2013.

[76] P.-J. Shih, C.-H. Lee, P.-C. Yeh, and K.-C. Chen, "Channel Codes for Reliability Enhancement in Molecular Communication," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 857–867, December 2013.

[77] B. Atakan, S. Galmes, and O. Akan, "Nanoscale Communication With Molecular Arrays in Nanonetworks," *IEEE Trans. NanoBioscience*, vol. 11, no. 2, pp. 149–160, June 2012.

[78] R. Qiu, Z. Hu, Z. Chen, N. Guo, R. Ranganathan, S. Hou, and G. Zheng, "Cognitive Radio Network for the Smart Grid: Experimental System Architecture, Control Algorithms, Security, and Microgrid Testbed," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 724–740, 2011.

[79] A. Ghassemi, S. Bavarian, and L. Lampe, "Cognitive Radio for Smart Grid Communications," in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2010, pp. 297–302.

[80] R. Deng, S. Maharjan, X. Cao, J. Chen, Y. Zhang, and S. Gjessing, "Sensing-Delay Tradeoff for Communication in Cognitive Radio Enabled Smart Grid," in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 155–160.

[81] R. Deng, J. Chen, X. Cao, Y. Zhang, S. Maharjan, and S. Gjessing, "Sensing-Performance Tradeoff in Cognitive Radio Enabled Smart Grid," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 302–310, 2013.

[82] M. Brew, F. Darbari, L. Crockett, M. B. Waddell, M. Fitch, S. Weiss, and R. Stewart, "UHF White Space Network for Rural Smart Grid Communications," in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 138–142.

[83] J. Huang, H. Wang, Y. Qian, and C. Wang, "Priority-Based Traffic Scheduling and Utility Optimization for Cognitive Radio Communication Infrastructure-Based Smart Grid," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 78–86, 2013.

[84] Z. Fadlullah, M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward Intelligent Machine-to-Machine Communications in Smart Grid," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 60–65, 2011.

[85] C.-J. M. Liang, N. B. Priyantha, J. Liu, and A. Terzis, "Surviving Wi-fi Interference in Low Power ZigBee Networks," in *ACM Conference on Embedded Networked Sensor Systems (SenSys)*, 2010, pp. 309–322.

[86] S.-C. Tuan, J.-C. Chen, H. T. Chou, and H.-H. Chou, "Optimization of Propagation Models for the Radio Performance Evaluation of Wireless Local Area Network," in *IEEE Ant. and Prop. Society Intl. Sym.*, 2003, pp. 146–149.

[87] M. Erol-Kantarci and H. T. Mouftah, "Wireless Multimedia Sensor and Actor Networks for the Next Generation Power Grid," *J. Ad-Hoc Networks*, vol. 9, no. 4, pp. 542 – 551, 2011.

[88] A. Bicen, O. Akan, and V. Gungor, "Spectrum-Aware and Cognitive Sensor Networks for Smart Grid Applications," *IEEE Commun. Mag.*, vol. 50, no. 5, pp. 158–165, 2012.

[89] D. Goodman, R. Valenzuela, K. T. Gayliard, and B. Ramamurthi, "Packet Reservation Multiple Access for Local Wireless Communications," *IEEE Trans. Commun.*, vol. 37, no. 8, pp. 885–890, 1989.

[90] S. Nanda, D. Goodman, and U. Timor, "Performance of PRMA: A Packet Voice Protocol for Cellular Systems," *IEEE Trans. Veh. Tech.*, vol. 40, no. 3, pp. 584–598, 1991.

[91] S. Nanda, "Analysis of Packet Reservation Multiple Access: Voice Data Integration for Wireless Networks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 1990, pp. 1984–1988.

[92] H. Qi and R. Wyrwas, "Markov Analysis for PRMA Performance Study," in *IEEE*

Vehicular Technology Conference (VTC), 1994, pp. 1184–1188.

[93] L. Jalloul, S. Nanda, and D. Goodman, "Packet Reservation Multiple Access over Slow and Fast Fading Channels," in *IEEE Vehicular Technology Conference (VTC)*, 1990, pp. 354–359.

[94] D. Park, J. Bang, and S. Tekinay, "Handoff Effect on Packet Reservation Multiple Access (PRMA) in Microcellular System," *IEEE Trans. Veh. Tech.*, vol. 52, no. 5, pp. 1232–1241, 2003.

[95] A. Brand and A. Aghvami, "Multidimensional PRMA with Prioritized Bayesian Broadcast - A MAC Strategy for Multiservice Traffic over UMTS," *IEEE Trans. Veh. Tech.*, vol. 47, no. 4, pp. 1148–1161, 1998.

[96] R. Fantacci, T. Pecorella, and I. Habib, "Proposal and Performance Evaluation of an Efficient Multiple Access Protocol for LEO Satellite Packet Networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 3, pp. 538–545, 2004.

[97] L. Hanzo, J. C. S. Cheung, R. Steele, and W. T. Webb, "A Packet Reservation Multiple Access assisted Cordless Telecommunication Scheme," *IEEE Trans. Veh. Tech.*, vol. 43, no. 2, pp. 234–244, 1994.

[98] A. Ghasemi and E. Sousa, "Optimization of Spectrum Sensing for Opportunistic Spectrum Access in Cognitive Radio Networks," in *IEEE Consumer Communications and Networking Conference (CCNC)*, 2007, pp. 1022–1026.

[99] H. Urkowitz, "Energy Detection of Unknown Deterministic Signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[100] W.-Y. Lee and I. Akyildiz, "Optimal Spectrum Sensing Framework for Cognitive Radio Networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3845–3857, 2008.

[101] Y.-C. Liang, Y. Zeng, E. Peh, and A. T. Hoang, "Sensing-Throughput Tradeoff for Cognitive Radio Networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, 2008.

[102] Y. Pei, Y.-C. Liang, K. Teh, and K. H. Li, "Sensing-Throughput Tradeoff for Cognitive Radio Networks: A Multiple-channel Scenario," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2009, pp. 1257–1261.

[103] Y. Pei, A. T. Hoang, and Y.-C. Liang, "Sensing-Throughput Tradeoff in Cognitive Radio Networks: How Frequently Should Spectrum Sensing be Carried Out?" in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2007, pp. 1–5.

[104] L. Tang, Y. Chen, E. Hines, and M.-S. Alouini, "Effect of Primary User Traffic on Sensing-Throughput Tradeoff for Cognitive Radios," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1063–1068, 2011.

[105] J. Zhang, H. Zhu, and H. Zhi, "On Channels Activity of Opportunistic Spectrum Sharing with Homogeneous Primary Users," in *IEEE International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, 2010, pp. 1–5.

[106] J. Zhang, L. Qi, and H. Zhu, "Optimization of MAC Frame Structure for Opportunistic Spectrum Access," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2036–2045, 2012.

[107] S.-Y. Lien, C.-C. Tseng, and K.-C. Chen, "Carrier Sensing Based Multiple Access Protocols for Cognitive Radio Networks," in *IEEE International Conference on*

Communications (ICC), 2008, pp. 3208–3214.

[108] C. Cordeiro and K. Challapali, "C-MAC: A Cognitive MAC Protocol for Multi-Channel Wireless Networks," in *IEEE International Conference on Dynamic Spectrum Access Networks (DySPAN)*, 2007, pp. 147–157.

[109] H. Li and R. Qiu, "Need-Based Communication for Smart Grid: When to Inquire Power Price?" in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2010, pp. 1–5.

[110] J. Frame, "Locational Marginal Pricing," in *IEEE Power Engineering Society (PES) Winter Meeting*, vol. 1, 2001, pp. 377–382.

[111] H. Song, C.-C. Liu, J. Lawarree, and R. Dahlgren, "Optimal Electricity Supply Bidding by Markov Decision Process," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 618–624, 2000.

[112] W. Yu and G. B. Sheblé, "Modeling Electricity Markets with Hidden Markov Model," *Elect. Power Syst. Research*, vol. 76, no. 6–7, pp. 445 – 451, 2006.

[113] H. Li, S. Gong, L. Lai, Z. Han, R. Qiu, and D. Yang, "Efficient and Secure Wireless Communications for Advanced Metering Infrastructure in Smart Grids," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1540–1551, 2012.

[114] H. Li, L. Lai, and R. Qiu, "Scheduling of Wireless Metering for Power Market Pricing in Smart Grid," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1611–1620, 2012.

[115] V. Gungor, B. Lu, and G. Hancke, "Opportunities and Challenges of Wireless Sensor Networks in Smart Grid," *IEEE Trans. Indust. Electronics*, vol. 57, no. 10, pp. 3557–3564, Oct 2010.

[116] J. Hill and D. Culler, "Mica: A Wireless Platform for Deeply Embedded Networks," *IEEE Micro*, vol. 22, no. 6, pp. 12–24, Nov 2002.

[117] T. Watteyne, A. Bachir, M. Dohler, D. Barthel, and I. Auge-Blum, "1-hopMAC: An Energy-Efficient MAC Protocol for Avoiding 1-hop Neighborhood Knowledge," in *IEEE International Conference on Sensor and Ad Hoc Communications and Networks (SECON)*, Sept 2006, pp. 639–644.

[118] K. Chowdhury, M. Di Felice, and I. Akyildiz, "TP-CRAHN: a Transport Protocol for Cognitive Radio Ad-Hoc Networks," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2009, pp. 2482–2490.

[119] D. S. J. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A High-Throughput Path Metric for Multi-hop Wireless Routing," in *ACM International Conference on Mobile Computing and Networking (MOBICOM)*, 2003, pp. 134–146.

[120] C.-T. Huang and J. Chang, "Responding to Security Issues in WiMAX Networks," *IEEE IT Professional*, vol. 10, no. 5, pp. 15–21, Sept 2008.

[121] "HF Interference, Procedures, and Tools," North AtlanticC Treaty Organization (NATO), TR IST-050, Jun. 2007. [Online]. Available: http://ftp.rta.nato.int/public/PubFullText/RTO/TR/RTO-TR-IST-050/$$TR-IST-050-ALL.pdf

[122] P. Kulkarni, S. Gormus, Z. Fan, and F. Ramos, "AMI Mesh Networks: A Practical Solution and Its Performance Evaluation," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1469–1481, 2012.

[123] E. Ancillotti, R. Bruno, and M. Conti, "The Role of the RPL Routing Protocol for Smart Grid Communications," *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 75–83, 2013.

[124] D. Wang, Z. Tao, J. Zhang, and A. Abouzeid, "RPL Based Routing for Advanced

Metering Infrastructure in Smart Grid," in *IEEE International Conference on Communications (ICC) Workshops, 2010*, 2010, pp. 1–6.

[125] P. Thulasiraman, "RPL Routing for Multigateway AMI Networks under Interference Constraints," in *IEEE International Conference on Communications (ICC)*, 2013, pp. 4477–4482.

[126] E. Ancillotti, R. Bruno, and M. Conti, "RPL Routing Protocol in Advanced Metering Infrastructures: An Analysis of the Unreliability Problems," in *IEEE/IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT), 2012*, 2012, pp. 1–10.

[127] P. Levis, T. Clausen, J. Hui, O. Gnawali, and J. Ko, "The Trickle Algorithm," Internet Engineering Task Force, RFC 6206, Mar. 2011. [Online]. Available: http://www.rfc-editor.org/rfc/rfc6206.txt

[128] T. Clausen, U. Herberg, and M. Philipp, "A Critical Evaluation of the IPv6 Routing Protocol for Low Power and Lossy Networks (RPL)," in *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2011, pp. 365–372.

[129] K. Chowdhury and I. Akyildiz, "CRP: A Routing Protocol for Cognitive Radio Ad Hoc Networks," *J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 794–804, 2011.

[130] S. Biswas and R. Morris, "ExOR: Opportunistic Multi-hop Routing for Wireless Networks," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 4, pp. 133–144, Aug. 2005.

[131] G. Schaefer, F. Ingelrest, and M. Vetterli, "Potentials of Opportunistic Routing in Energy-Constrained Wireless Sensor Networks," in *European Conference on Wireless Sensor Networks (EWSN)*, 2009, pp. 118–133.

[132] H. Myung, J. Lim, and D. Goodman, "Single Carrier FDMA for Uplink Wireless Transmission," *IEEE Veh. Tech. Mag.*, vol. 1, no. 3, pp. 30–38, 2006.

[133] G. Berardinelli, L. Ruiz de Temino, S. Frattasi, M. Rahman, and P. Mogensen, "OFDMA vs. SC-FDMA: Performance Comparison in Local Area IMT-A Scenarios," *IEEE Wireless Commun.*, vol. 15, no. 5, pp. 64–72, October 2008.

[134] 3GPP, "Physical Layer Aspects for Evolved UTRA (Release 7)," 3rd Generation Partnership Project (3GPP), TR 25.814, May 2006. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/25814.htm

[135] I. Wong, O. Oteri, and W. Mccoy, "Optimal resource allocation in uplink sc-fdma systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2161–2165, 2009.

[136] O. Delgado and B. Jaumard, "Scheduling and Resource Allocation in LTE Uplink with a Delay Requirement," in *IEEE Communication Networks and Services Research Conference (CNSR)*, May 2010, pp. 268–275.

[137] M. Wang, Z. Zhong, and Q. Liu, "Resource Allocation for SC-FDMA in LTE Uplink," in *IEEE International Conference on Service Operations, Logistics, and Informatics (SOLI)*, July 2011, pp. 601–604.

[138] K. Zheng, F. Hu, W. Wang, W. Xiang, and M. Dohler, "Radio Resource Allocation in LTE-Advanced Cellular Networks with M2M Communications," *IEEE Commun. Mag.*, vol. 50, no. 7, pp. 184–192, July 2012.

[139] A. Ahmad and M. Assaad, "Power Efficient Resource Allocation in Uplink SC-FDMA Systems," in *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept 2011, pp. 1351–1355.

[140] D. Dechene and A. Shami, "Energy Efficient Resource Allocation in SC-FDMA Uplink with Synchronous HARQ Constraints," in *IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–5.

[141] A. Aijaz and A.-H. Aghvami, "On Radio Resource Allocation in LTE Networks with Machine-to-Machine communications," in *IEEE Vehicular Technology Conference (VTC Spring)*, June 2013, pp. 1–5.

[142] V. Rodoplu and T. Meng, "Bits-per-Joule Capacity of Energy-Limited Wireless Networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 857–865, March 2007.

[143] D. Wu and R. Negi, "Effective Capacity: A Wireless Link Model for Support of Quality of Service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, 2003.

[144] D. Y. Gao, "Canonical Dual Transformation Method and Generalized Triality Theory in Nonsmooth Global Optimization," *J. Global Opt.*, vol. 17, no. 1-4, 2000.

[145] R. Giuliano and F. Mazzenga, "Exponential Effective SINR Approximations for OFDM/OFDMA-based Cellular System Planning," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4434–4439, 2009.

[146] J. Tang and X. Zhang, "Quality-of-Service Driven Power and Rate Adaptation over Wireless Links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3058–3068, 2007.

[147] F. Calabrese, "Scheduling and Link Adaptation for Uplink SC-FDMA Systems," *Ph.D. dissertation, Aalborg University*, 2009.

[148] J. Tang and X. Zhang, "Cross-Layer-Model Based Adaptive Resource Allocation for Statistical QoS Guarantees in Mobile Wireless Networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2318–2328, June 2008.

[149] ——, "Power-Delay Tradeoff over Wireless Networks," in *International Symposium on World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2008, pp. 1–12.

[150] D. Y. Gao, R. lin Sheu, S.-Y. Wu, and C. K. L. Teo, "Canonical Dual Approach for Solving 0-1 Quadratic Programming Problems," *J. Industrial and Management Optimization*, vol. 4, pp. 125–142, 2007.

[151] A. Ahmad and M. Assaad, "Polynomial-Complexity Optimal Resource Allocation Framework for Uplink SC-FDMA Systems," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2011, pp. 1–5.

[152] ——, "Canonical Dual Method for Resource Allocation and Adaptive Modulation in Uplink SC-FDMA System," *arXiv:1103.4547*, 2012.

[153] D. Gao, N. Ruan, and H. Sherali, "Canonical Dual Solutions for Fixed Cost Quadratic Programs," in *Optimization and Optimal Control*. Springer, 2010, pp. 139–156.

[154] E. Pourjafari and H. Mojallali, "Solving Nonlinear Equations Systems with a New Approach based on Invasive Weed Optimization Algorithm and Clustering ," *J. Swarm and Evo. Comput.*, vol. 4, no. 0, pp. 33 – 43, 2012.

[155] 3GPP, "Further Advancements for E-UTRA Physical Layer Aspects," 3rd Generation Partnership Project (3GPP), TR 36.814, 2010. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36814.htm

[156] A. Aijaz, M. Nakhai, and A.-H. Aghvami, "Power Efficient Uplink Resource Allocation in LTE Networks under Delay QoS Constraints," in *IEEE Global Communications Conference (GLOBECOM)*, 2014, pp. 1–6, (accepted for publication).

[157] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the Air: Practical Wireless Network Coding," *IEEE/ACM Trans. Networking,*, vol. 16,

no. 3, pp. 497–510, June 2008.

[158] L. Ma, Z. Lin, Z. Zhang, G. Mao, and B. Vucetic, "Improving Reliability in Lossy Wireless Networks Using Network Coding," in *IEEE International Conference on Communications (ICC)*, June 2013, pp. 312–316.

[159] T. Nakano and J.-Q. Liu, "Design and Analysis of Molecular Relay Channels: An Information Theoretic Approach," *IEEE Trans. NanoBioscience*, vol. 9, no. 3, pp. 213–221, Sept 2010.

[160] A. Einolghozati, M. Sardari, and F. Fekri, "Relaying in Diffusion-based Molecular Communication," in *IEEE International Symposium on Information Theory (ISIT)*, July 2013, pp. 1844–1848.

[161] B. Unluturk, D. Malak, and O. Akan, "Rate-Delay Tradeoff With Network Coding in Molecular Nanonetworks," *IEEE Trans. Nanotechnology*, vol. 12, no. 2, pp. 120–128, March 2013.

[162] I. Llatser, A. Cabellos-Aparicio, M. Pierobon, and E. Alarcon, "Detection Techniques for Diffusion-based Molecular Communication," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 12, pp. 726–734, December 2013.

[163] J. Philibert, "One and a Half Century of Diffusion: Fick, Einstein, Before and Beyond," *Diffusion Fundamentals*, vol. 2, pp. 1–10, 2005.

[164] M. Pierobon and I. Akyildiz, "Diffusion-Based Noise Analysis for Molecular Communication in Nanonetworks," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2532–2547, June 2011.

[165] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 2002.

[166] M. Şükrü Kuran, H. B. Yilmaz, T. Tugcu, and B. Özerman, "Energy Model for Communication via Diffusion in Nanonetworks," *J. Nano Commun. Networks*, vol. 1, no. 2, pp. 86 – 95, 2010.

[167] S. Kadloor, R. Adve, and A. Eckford, "Molecular Communication Using Brownian Motion With Drift," *IEEE Trans. NanoBioscience*, vol. 11, no. 2, pp. 89–99, June 2012.

[168] G. Alfano and D. Miorandi, "On Information Transmission Among Nanomachines," in *IEEE International Conference on Nano-Networks (NanoNet) Workshops*, Sept 2006, pp. 1–5.

[169] A. Credi, V. Balzani, S. J. Langford, and J. F. Stoddart, "Logic Operations at the Molecular Level. An XOR Gate Based on a Molecular Machine," *J. American Chem. Soc.*, vol. 119, no. 11, pp. 2679–2681, 1997.

[170] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative Spectrum Sensing in Cognitive Radio Networks: A Survey ," *J. Phy. Commun.*, vol. 4, no. 1, pp. 40 – 62, 2011.

[171] CMCC TSG R2-113197, "Performance Comparison of Access Class Barring and MTC Specific Backoff Schemes for MTC," 3rd Generation Partnership Project (3GPP) Meeting, Tech. Rep., Aug. 2010.

[172] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro, and K. Hugl, "Device-to-Device Communication as an Underlay to LTE-Advanced Networks," *IEEE Commun. Mag.*, vol. 47, no. 12, pp. 42–49, Dec 2009.

[173] J. Zhang, L. Shan, H. Hu, and Y. Yang, "Mobile Cellular Networks and Wireless Sensor

Networks: Toward Convergence," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 164–169, March 2012.

[174] C.-S. Chang, "Stability, Queue Length, and Delay of Deterministic and Stochastic Queueing Networks," *IEEE Trans. Automatic Control*, vol. 39, no. 5, pp. 913–931, May 1994.

[175] F. Kelly, S. Zachary, and I. Ziedins, *Stochastic Networks: Theory and Applications*. Oxford University Press, 1996, vol. 4.